Shaun Gallagher
Daniel Schmicking

Editors

Handbook of Phenomenology and Cognitive Science

Springer

Handbook of Phenomenology
and Cognitive Science

Shaun Gallagher · Daniel Schmicking
Editors

Handbook of Phenomenology and Cognitive Science

Editors
Shaun Gallagher
University of Central Florida
Orlando FL
USA
gallaghr@mail.ucf.edu

Daniel Schmicking
Johannes-Gutenberg-Universität Mainz
Mainz, Germany
schmicki@uni-mainz.de

Printed on acid-free paper

Preface

This volume explores the essential issues involved in bringing phenomenology together with the cognitive sciences, and provides some examples of research located at the intersection of these disciplines. The topics addressed here cover a lot of ground, including questions about naturalizing phenomenology, the precise methods of phenomenology and how they can be used in the empirical cognitive sciences, specific analyses of perception, attention, emotion, imagination, embodied movement, action and agency, representation and cognition, intersubjectivity, language and metaphor. In addition there are chapters that focus on empirical experiments involving psychophysics, perception, and neuro- and psychopathologies.

The idea that phenomenology, understood as a philosophical approach taken by thinkers like Husserl, Heidegger, Sartre, Merleau-Ponty, and others, can offer a positive contribution to the cognitive sciences is a relatively recent idea. Prior to the 1990s, phenomenology was employed in a critique of the first wave of cognitivist and computational approaches to the mind (see Dreyfus 1972). What some consider a second wave in cognitive science, with emphasis on connectionism and neuroscience, opened up possibilities for phenomenological intervention in a more positive way, resulting in proposals like neurophenomenology (Varela 1996). Thus, brain-imaging technologies can turn to phenomenological insights to guide experimentation (see, e.g., Jack and Roepstorff 2003; Gallagher and Zahavi 2008). But even more important, phenomenology has played a significant role in initiating a third wave that considers the cognitive system to include not just the brain, but the body as a whole, situated in physical and social environments. This involves a shift that now emphasizes embodied cognition, enactive perception, and dynamical systems, and integrates the work of phenomenologists (see Gallagher and Varela 2003; Varela et al. 1991; Thompson 2007). It was this shift that motivated the founding of the journal *Phenomenology and the Cognitive Sciences*, and that spawned a

significant number of academic conferences in both Europe and the United States.[1] This volume does not try to trace this history. It is more properly concerned with laying out the current state of affairs with respect to the various topics that fall into this area of research. We think the papers collected here speak for themselves, and speak clearly about issues important for both philosophers and scientists who study human experience and cognition.

We thank the Siemens Corporation for funding a Visiting Scholar position in German Philosophy at the University of Central Florida in Fall 2007, and for supporting a conference on *Phenomenology and the Science of Consciousness* in March 2008. The first allowed a research visit by Dan Schmicking; the second supported a conference that brought several of this volume's contributors together to share their work. Thanks also to Leslie Gale and Jaci Schock at UCF's Philosophy Department for their continuing administrative support.

<div align="right">Dan Schmicking and Shaun Gallagher</div>

References

Dreyfus H (1972) What computers can't do: a critique of artificial reason. MIT Press, Cambridge, MA

Gallagher S, Varela F (2003) Redrawing the map and resetting the time: phenomenology and the cognitive sciences. Can J Philos Suppl 29:93–132

Gallagher S, Zahavi D (2008) The phenomenological mind. Routledge, London

Jack A, Roepstorff A (2003) Trusting the subject, vol 1. Imprint Academic, Exeter

Thompson E (2007) Mind in life: biology, phenomenology, and the sciences of mind. Harvard University Press, Cambridge, MA

Varela F, Thompson E, Rosch E (1991) The embodied mind: cognitive science and human experience. MIT Press, Cambridge, MA

Varela F (1996) Neurophenomeology. J Conscious Stud 3:330–349[1]

---

[1]For example, *Cognition: Embodied, Embedded, En active, Extended*, University of Central Florida (October 2007); *Situated Cognition: Perspectives from Phenomenology and Science*, Durham University, UK (August 2006); *Phenomenology, Intersubjectivity, and Theory of Mind*. University of Central Florida, Orlando (January 2005); *The Embodied Mind*. Copenhagen (June 2004); *Intersubjectivity and Embodiment. Perspectives from Phenomenology and the Cognitive Sciences*. Louvain (September 2003); *The Imagination in Phenomenology and Cognitive Science*. London (July 2002); *Intentionality and Experience*. Copenhagen (June 2001); *Beyond the Hard Problem: Consequences of Neurophenomenology*. Boulder, Colorado (May 25–26, 2001); *Phenomenological and Experimental Approaches to Cognition*. CREA, Paris (June, 2000).

Contents

Part III Embodiment

Part IV Intersubjectivity

Part V Perception, Action and Enactive Phenomenology

Part VI Language and Meaning

Part I
Phenomenology and Experimental Cognitive Science

Naturalized Phenomenology

Dan Zahavi

It is always risky to make sweeping statements about the development of philosophy, but if one were nevertheless asked to describe twentieth century philosophy in broad strokes, one noteworthy feature might be the following: Whereas important figures at the beginning of the century, figures such as Frege and Husserl, were very explicit in their rejection of naturalism (both are known for their rejection of the attempt to naturalize the laws of logic, that is, for their criticism of psychologism), the situation has changed considerably. Today many philosophers – not the least within analytical philosophy – would subscribe to some form of naturalism. In fact, naturalism is seen by many as the default metaphysical position. If you don't subscribe to naturalism you must be subscribing to some form of Cartesian substance dualism. Thus, whereas 20 or 30 years ago one might have been inclined to characterize the development of twentieth century philosophy in terms of a linguistic turn, a turn from a philosophy of subjectivity to a philosophy of language, it might today be more apt to describe the development in terms of a turn from anti-naturalism to naturalism.

What are the implications of this turn? It has some rather decisive metaphilosophical implications, that is, it has implications for the way we view the relation between philosophy and positive science. According to some readings, a commitment to naturalism simply amounts to taking one's departure in what is natural (rather than supernatural), but I think it is fair to say that the use of the term in the current discourse mainly signals an orientation towards natural science. As Sellars famously put it, "in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is and of what is not that it is not" (Sellars 1963, 173). However, insofar as naturalists would consider the scientific account of reality authoritative, a commitment to naturalism is bound to put pressure on the idea that philosophy (including phenomenology) can make a distinct and autonomous contribution to the study of reality.

D. Zahavi
Department of Media, Cognition and Communication & Center for Subjectivity Research, University of Copenhagen, Copenhagen, Denmark

A vivid illustration of this tension can be found in the field of consciousness studies. As Francis Crick insists "it is hopeless to try to solve the problems of consciousness by general philosophical arguments; what is needed are suggestions for new experiments that might throw light on these problems." (Crick 1995, 19). Indeed, on Crick's view, "the study of consciousness is a scientific problem. […] There is no justification for the view that only philosophers can deal with it." (Crick 1995, 258). Quite on the contrary in fact, since philosophers "have had such a poor record over the last 2,000 years that they would do better to show a certain modesty rather than the lofty superiority that they usually display." (Crick 1995, 258). This is not to say that philosophers cannot make some kind of contribution, but they must "learn how to abandon their pet theories when the scientific evidence goes against them or they will only expose themselves to ridicule."(Crick 1995, 258). In short, philosophers are welcome to join the common enterprise, but only as junior partners. Indeed, one suspects that philosophy (of mind) on Crick's view will ultimately turn out to be dispensable. Whatever contribution it can make is propaedeutical and will eventually be replaced by a proper scientific account.

It is doubtful whether all philosophers committed to naturalism would accept this line of thought, but there is no question that naturalism does pose a challenge to philosophy. In the following, my focus will be on phenomenology. How should phenomenology respond to this challenge? What sense can we make of recent proposals to naturalize phenomenology? A correct appraisal is, however, complicated by the fact that naturalism is far from being a univocal position. I will take my point of departure in a rather classical reductionist conception of naturalism, not only because this was the kind of position that Husserl was originally reacting against, but also because it still remains a widespread and influential view. Whether this ultimately amounts to a satisfying conception of naturalism is a question I will return to later.

Husserl's Anti-naturalism

To start with, let us briefly recapitulate the character and motivation for Husserl's anti-naturalism. In the long essay *Philosophy as Rigorous Science*, Husserl describes naturalism as a fundamentally flawed philosophy (Husserl 1987, 41) and argues that it has typically had two different aims: the naturalization of ideality and normativity, and the naturalization of consciousness (Husserl 1987, 9). In his view, however, both attempts fail and both are misguided. The naturalistic reduction of ideality leads to scepticism (Husserl 1987, 7; 1984, 47). This, in fact, was one of Husserl's main arguments in his famous fight against psychologism in the *Logical Investigations*. As for Husserl's criticism of the attempt to naturalize consciousness, he explicitly contrasts his own phenomenology of consciousness with a natural scientific account of consciousness (Husserl 1987, 17). Both disciplines investigate consciousness, but according to Husserl they do so in utterly different manners. And to suggest that the phenomenological account could be absorbed, or reduced,

or replaced by a naturalistic account is for Husserl sheer nonsense. This is not to say that Husserl doesn't respect natural science, but as he famously put it in *Ideas I*, "When it is actually natural science that speaks, we listen gladly and as disciples. But it is not always natural science that speaks when natural scientists are speaking; and it assuredly is *not* when they are talking about 'philosophy of Nature' and 'epistemology as a natural science'" (Husserl 1982, 39).

But why does Husserl oppose the attempt to implement a thorough naturalistic account of consciousness? Because naturalism in his view is incapable of doing full justice to consciousness. Not only has it – in the shape of experimental psychology – lost sight of (subjective) consciousness (Husserl 1987, 104), but even more importantly, naturalism treats consciousness as a real occurrent entity in the world, that is, it conceives of consciousness as an object in the world, on a par with – though possibly more complex than – volcanoes, waterfalls, ice crystals, gold nuggets, rhododendrons or black holes. But on Husserl's view this is unacceptable since consciousness rather than merely being an object in the world, is also a subject for the world, that is, a necessary condition of possibility for any entity to appear as an object in the way it does and with the meaning it has. To put it differently, according to Husserl, the decisive limitation of naturalism is that it is incapable of recognizing the *transcendental* dimension of consciousness.

One way to interpret Kant's revolutionary *Copernican turn* is by seeing it as amounting to the realization that our cognitive apprehension of reality is more than a mere mirroring of a preexisting world. Moreover, Kant transformed the pre-critical search for the most fundamental building blocks of reality into a reflection on what conditions something must satisfy in order to count as "real"; what is the condition of possibility for the appearance of objects? With various modifications this idea was picked up by Husserl and subsequent phenomenologists. Indeed, the reason why phenomenologists have emphasized the importance of the first-person perspective and investigated the fundamental structures of consciousness and selfhood in great detail has not been because they considered such an investigation a goal in itself – if so, phenomenology would have remained a form of philosophical psychology or philosophical anthropology – rather the analysis was motivated by transcendental philosophical considerations.

Naturalism is typically distinguished by methodological as well as metaphysical commitments. The methodological commitment amounts to the idea that the right criteria for justification are those found in and employed by the natural sciences. The metaphysical commitment amounts to the monistic view that there is only one kind of thing, namely things with natural properties, so that everything existing is natural. Jointly, the two commitments amounts to the view that everything (including everything pertaining to human life, such as consciousness, culture and history) has to be studied by the methods of natural science (cf. Aikin 2006, 318). Moreover, this outlook is frequently tied to an explicit endorsement of metaphysical realism. One way to define metaphysical realism is to see it as being guided by a certain conception of knowledge. Knowledge is taken to consist in a faithful mirroring of a mind-independent reality. It is taken to be knowledge of a reality which exists independently of that knowledge, and indeed independently of any thought and

experience (Williams 2005, 48). If we want to know true reality, we should aim at describing the way the world is, independently of all the ways in which it happens to present itself to us human beings, that is, we should aim for a description where all traces of ourselves have been removed. Metaphysical realism assumes that everyday experience combines subjective and objective features and that we can reach an objective picture of what the world is really like by stripping away the subjective. It consequently argues that there is a clear distinction to be drawn between the properties things have 'in themselves' and the properties which are 'projected by us'. Whereas the world of appearance, the world as it is for us in daily life, combines subjective and objective features, science captures the objective world, the world as it is in itself.

This assumption has been questioned by phenomenologists. They have criticized the suggestion that science can provide us with a description from a view from nowhere as if science simply mirrors the way in which pre-existing and mind-independent nature classifies itself. They have argued that a view from nowhere is unattainable, just as they would deny that it is possible to look at our experiences sideways on to see whether they match with reality. This is so, not because such views are incredibly hard to reach, but because the very idea of such views is nonsensical.

For phenomenology, science is not simply a collection of systematically inter-related justified propositions. Science is performed by somebody; it is a specific theoretical stance towards the world. This stance did not fall down from the sky; it has its own presuppositions and origins. Scientific objectivity is something to strive for, but it rests on the observations and experiences of individuals; it is knowledge shared by a community of experiencing subjects and presupposes a triangulation of points of view or perspectives. Thus, according to this view, rather than being as such a hindrance or obstacle, consciousness turns out to be a far more important requisite for objectivity and the pursuit of scientific knowledge than, say, micro-scopes and scanners.

For Husserl, naturalism takes its subject matter, nature, for granted. Reality is assumed to be out there, waiting to be discovered and investigated. And the aim is then to acquire a strict and objectively valid knowledge about this given realm. But for Husserl this attitude must be contrasted with the properly philosophical attitude, which critically questions the very foundation of experience and scientific thought (Husserl 1987, 13–14). Philosophy is a discipline which doesn't simply contribute to or extend the scope of our scientific knowledge, but which instead investigates the basis of this knowledge and asks how it is possible. Naturalism has denied the existence of a particular philosophical method, and has claimed that philosophy should employ the same method that all strict sciences are using, the natural scientific method. But for Husserl this line of reasoning merely displays that one has failed to understand what philosophy is all about. Philosophy has its own aims and methodological requirements; requirements that for Husserl are epitomized in his notion of *phenomenological reduction* (Husserl 1984, 238–239). For Husserl, the reduction is meant to make us maintain the radical difference between philo-sophical reflection and all other modes of thought. As he wrote in 1907: "*Thus, the*

'phenomenological reduction' is simply the requirement always to abide by the sense of the proper investigation, and not to confuse epistemology with a natural scientific (objectivistic) investigation" (Husserl 1984, 410). Every positive science rests upon a field of givenness or evidence that is presupposed but not investigated by the sciences themselves. In order to make this dimension accessible, a new type of inquiry is called for, a type of inquiry which "precedes all natural knowledge and science and points in a quite different direction than natural science" (Husserl 1984, 176). This, of course, is one reason why the phenomenological attitude has frequently been described as an unnatural direction of thought (Husserl 2001, I. 170). But to describe phenomenology as unnatural is of course also to deny any straightforward continuity between philosophy and natural science.

Husserl's view on this issue seems pretty much to have been shared by Heidegger. In a famous talk entitled *Phänomenologie und Theologie* written in 1927 Heidegger argued that within the different positive sciences we can speak of relative differences. One science, say anthropology, investigates one specific realm, another science, say biology, investigates another realm. Between the positive sciences and phenomenological philosophy there is also a difference, but this difference is not a relative, but an absolute one. For whereas the positive sciences are ontical sciences which are interested in beings (*das Seiende*), phenomenology is an ontological science which is concerned with Being (*Sein*). It is in this context that Heidegger famously remarks that there are more similarities between theology and chemistry (both of which are concerned with beings) than between theology and philosophy (Heidegger 1978, 49). Given this outlook we can at most expect a kind of one-way communication between philosophy and science, where the former constrains the latter.

Transcendental Philosophy and Philosophical Psychology

So far, phenomenology's response to naturalism is rather unequivocal. Contrary to some proposals, it is not naturalism's classical endorsement of some form of reductive materialism that constitutes the main obstacle to a reconciliation. It is not as if matters would improve if naturalism opted for some version of emergentism or property dualism. The real problem has to do with naturalism's commitment to scientism and metaphysical realism. Indeed given such commitments, the attempt to naturalize phenomenology seems fundamentally misguided. As Husserl explained to the Neo-Kantian Rickert in a letter from 1915, he considered the fight against naturalism – a fight he had devoted his life to – indispensable for the progress of philosophy (Husserl 1994, V. 178). Indeed, for Husserl a phenomenologist who embraced naturalism would in effect have ceased being a philosopher.

To naturalize phenomenology might indeed – to use a formulation employed by the editors of the volume *Naturalizing Phenomenology* – be a question of integrating phenomenology into an explanatory framework where every acceptable property is made continuous with the properties admitted by natural science (Roy et al. 1999,

1–2). However, let us at this point reassess this initial definition of naturalization, and consider a rather different (and more modest) proposal. To naturalize phenomenology might simply be a question of letting phenomenology engage in a fruitful exchange and collaboration with empirical science. Phenomenology does study phenomena that are part of nature and therefore also open to empirical investigation, and insofar as phenomenology concerns itself with such phenomena it should be informed by the best available scientific knowledge. The phenomenological credo 'To the things themselves' calls for us to let our experience guide our theories. We should pay attention to the way in which we experience reality. Empirical scientists might not pay much attention to the formal structure of phenomenality, but as empirical researchers they do in fact pay quite a lot of attention to concrete phenomena, and might consequently be less apt to underestimate the richness, complexity and variety of phenomena than the average philosopher. As Merleau-Ponty puts it in *La Nature*:

"How thus not to be interested in science in order to know what Nature is? If Nature is something all-encompassing, we cannot think it starting from concepts, let alone deductions, rather we must think it starting from experience, and in particular, experience in its most regulated form – that is, science." (Merleau-Ponty 2000, 87 [transl. modified])

To put it differently, the ultimate aim of phenomenology is to provide a transcendental philosophical clarification, and as such its aim differs from that of empirical science. However, there is more to phenomenology than this ultimate goal. Phenomenology also offers detailed analyses of various aspects of consciousness, including perception, imagination, embodiment, memory, self-experience, temporality, etc. In offering such analyses, phenomenology addresses issues that are crucial for an understanding of the true complexity of consciousness and might even offer a conceptual framework for understanding the mind that is of considerably more value than some of the models currently in vogue in cognitive science. But for the very same reason, it should also be clear that phenomenology deals with topics that it shares with other disciplines, and it would be wrong to insist that it should simply ignore empirical findings pertaining to these very topics. Does this entail that a phenomenological account of perception or action should necessarily be informed and constrained by, say, investigations of the neuronal mechanisms and processes involved in action and perception? As I will argue in a moment, in some cases discoveries of the latter kind could motivate us to take another look at the phenomenology, in order to ascertain whether we got it right the first time, but generally speaking a phenomenological account of perception and action is an attempt to do justice to the first-person perspective, it seeks to understand the experience in terms of the meaning it has for the subject, and doesn't address the subpersonal mechanisms that might enable us to experience the way we do. However, we shouldn't overlook that disciplines such as psychopathology, neuropathology, developmental psychology, cognitive psychology, anthropology etc. can provide person-level *descriptions* that might be of phenomenological relevance. The examples are legion, but if one were to mention a few, one could single out (1) neuropsychological descriptions of anosognosic disorders of body-awareness, (2) psychopathological descriptions of schizophrenic disturbances of self-experience

and intentionality, (3) developmental descriptions of social interactions in early childhood, (4) ethnological descriptions of culture specific emotions, (5) descriptions of various types of memory provided by cognitive psychology, etc.

So on this proposal, the naturalization of phenomenology wouldn't merely consist in stressing the usefulness of phenomenological analyses and distinctions for, say, cognitive science. The point wouldn't merely be that phenomenology might prove indispensable if we wish to obtain a precise description of the explanandum – a *sine qua non* for any successful attempt to identify and localize the relevant neurobiological correlate. It wouldn't merely be a question of employing phenomenological insights in the empirical investigation of the mind. Rather, the idea would be that the influence goes both ways, that is, it would also be a question of letting phenomenology profit from – and be challenged by – empirical findings. This is why it is entirely appropriate to speak of a *mutual enlightenment* (cf. Gallagher 1997).

Various complementary proposals are currently in the offering, when it comes to cashing out this idea in more concrete details. One proposal entitled *neurophenomenology* was initially proposed by Varela (1996) and subsequently further developed by Lutz (2002), Lutz and Thompson (2003), and Thompson (2007). The basic idea here is to train the experimental subjects to gain greater intimacy with their own experiences. The subjects are subsequently asked to provide description of these experiences using an open-question format, and thus without the imposition of pre-determined theoretical categories. The ensuing descriptive categories are subsequently validated intersubjectively and then used to interpret correlated measurements of behavior and brain activity. At the same time, however, it is also suggested that, say, a consideration of insights from neurobiology and dynamical systems theory can help us improve and refine the classical phenomenological analyses (see Varela 1997). How is that supposed to happen? The basic idea is quite simple: Let us assume that our initial phenomenological description presents us with what appears to be a simple and unified phenomena. When studying the neural correlates of this phenomena, we discover that two quite distinct mechanisms are involved; mechanisms that are normally correlated with distinctive experiential phenomena, say, perception and memory. This discovery might motivate us to return to our initial phenomenological description in order to see whether the phenomenon in question is indeed as simple as we thought. Perhaps a more careful analysis will reveal that it harbors a concealed complexity. However, it is very important to emphasize that the discovery of a significant complexity on the sub-personal level – to stick to this simple example – cannot by itself force us to refine or revise our phenomenological description. It can only serve as motivation for further inquiry. Thus, it is certainly not being suggested that there is a straightforward isomorphism between the sub-personal and personal level. Ultimately, the only way to justify a claim concerning a complexity on the phenomenological level is by cashing it out in experiential terms.

More recently, Gallagher (2003) has made a slightly different proposal which he has entitled *front-loaded phenomenology*. Rather than focusing on the training of experimental subjects, the idea is here to start with the experimental design, and to

allow insights developed in phenomenological analyses to inform the way experiments are set up. To take a concrete example, let us consider the issue of self-consciousness. Within developmental psychology, the so-called mirror-recognition task has occasionally been heralded as the decisive test for self-consciousness. From around 18 months of age, children will engage in self-directed behavior when confronted with their mirror-image, and it has been argued that self-consciousness is only present from the moment the child is capable of recognizing itself in the mirror (cf. Lewis 2003). Needless to say, this line of reasoning makes use of a very specific notion of self-consciousness. Rather than simply letting phenomenological insights guide our interpretation of the results obtained through the testing of mirror-recognition, one possibility would be to let the phenomenological account and analysis of pre-reflective self-consciousness guide our design of the experimental paradigm. It would no longer involve the testing of mirror-recognition – which phenomenologists would typically consider evidence for the presence of a rather sophisticated form of self-consciousness – but, for instance, aim at detecting the presence of far more primitive forms of proprioceptive body-awareness. To front-load phenomenology, however, does not imply that one simply presupposes or accepts well rehearsed phenomenological results. Rather it involves testing those results and more generally it incorporates a dialectical movement between previous insights gained in phenomenology and preliminary trials that will specify or extend these insights for purposes of the particular experiment or empirical investigation (Gallagher 2003).

Are there any precedents in classical phenomenology for such integrative approaches? Let us briefly consider and compare the cases of Husserl and Merleau-Ponty, respectively. In several of his writings, Husserl distinguishes two different phenomenological approaches to consciousness. On the one hand, we have *transcendental phenomenology*, and on the other, we have what he calls *phenomenological psychology* (Husserl 1977). What is the difference between these two approaches? Both of them deal with consciousness, but they do so with rather different agendas in mind. For Husserl, the task of phenomenological psychology is to investigate intentional consciousness in a non-reductive manner, that is, in a manner that respects its peculiarity and distinctive features. Phenomenological psychology is a form of philosophical psychology which takes the first-person perspective seriously, but which – in contrast to transcendental phenomenology – remains within the natural attitude. The difference between the two is consequently that phenomenological psychology might be described as a regional-ontological analysis which investigates consciousness for its own sake. In contrast, transcendental phenomenology is a much more ambitious global enterprise. It is interested in the constitutive dimension of subjectivity, that is, it is interested in an investigation of consciousness in so far as consciousness is taken to be a condition of possibility for meaning, truth, validity, and appearance.

What is the relevance of this distinction? Although Husserl's primary aim was the development of transcendental phenomenology, he was not blind to the fact that his analyses might have ramifications for and be of pertinence to the psychological study of consciousness, and vice versa. As Husserl wrote: "every analysis or theory of

transcendental phenomenology – including […] the theory of the transcendental constitution of an objective world – can be carried out in the natural realm, when we give up the transcendental attitude. Eidetically and empirically, a *pure* psychology – a psychology that merely explicates what belongs to the psyche, to a concrete human Ego, as its own intentional essence – corresponds to a *transcendental phenomenology*, and vice versa" (Husserl 1999, 131 [transl. modified]). Husserl consequently spoke of a parallelism between phenomenological psychology and transcendental phenomenology and claimed that it is possible to step from one to the other through an attitudinal change. In fact, in *Cartesian Meditations* he even writes that it is pointless to treat transcendental phenomenology and the positive science of intentional psychology separately. At first he suggests that the former should pave the way, and that the latter could then take over some of the results (without having to bother with the transcendental considerations), but as he eventually goes on to say, in its core intentional psychology (the study of consciousness) contains a transcendental dimension, is part of transcendental philosophy, though this will remain concealed until psychology is relieved of its naivety (Husserl 1999, 147). Might such considerations allow for the possibility that empirical findings (if based on a meticulous analysis of the phenomena and if subjected to the requisite modifications) could be taken up by, and consequently influence or constrain the analysis of transcendental subjectivity? I see no reason why not. Consider for instance Husserl's discussion in *Ideas II* of the effects of consuming the anthelmintic drug santonin (Husserl 1989, 67–69), or to take a more fundamental example, consider the fact that Husserl's ontological way to the reduction takes a careful description of a specific ontological region as guiding-line for the subsequent transcendental analysis (Husserl 1970, 170–174).

If we move on to Merleau-Ponty, it is well known that he already in his first major work *The Structure of Behavior* discusses such diverse authors as Pavlov, Freud, Koffka, Piaget, Watson, and Wallon. The last sub-chapter of the book carries the heading "Is There Not a Truth in Naturalism?" It contains a criticism of Kantian transcendental philosophy, and on the very final page of the book, Merleau-Ponty calls for a redefinition of transcendental philosophy that makes it pay heed to the real world (Merleau-Ponty 1963, 224). Thus, rather than making us choose between either an external scientific explanation or an internal phenomenological reflection, a choice which would rip asunder the living relation between consciousness and nature, Merleau-Ponty asks us to reconsider the very opposition, and to search for a dimension that is beyond both objectivism and subjectivism.

This interest in positive science, in its significance for phenomenology, remains prominent in many of Merleau-Ponty's later works as well. His use of neuropathology (Gelb and Goldstein's famous Schneider-Case) in *Phenomenology of perception* is well known. For some time, in the years 1949–1952, Merleau-Ponty even held a chair in Child Psychology at the Sorbonne. As for his last writings, a representative statement is found in *Signs*, where Merleau-Ponty declares that "the ultimate task of phenomenology as philosophy of consciousness is to understand its relationship to non-phenomenology. What resists phenomenology within us – natural being, the 'barbarous' source Schelling spoke of – cannot remain outside phenomenology and should have its place within it" (Merleau-Ponty 1964, 178).

What is interesting and important is that Merleau-Ponty didn't conceive of the relation between transcendental phenomenology and empirical science as a question of how to apply already established phenomenological insights on empirical issues. It wasn't simply a question of how phenomenology might constrain positive science. On the contrary, Merleau-Ponty's idea was that phenomenology itself can be changed and modified through its dialogue with the empirical disciplines. In fact, it needs this confrontation if it is to develop in the right way.[1] And mind you, Merleau-Ponty held on to this view without thereby reducing phenomenology to merely yet another empirical science, without thereby dismissing its transcendental philosophical nature (Merleau-Ponty 1962, 63).[2]

In order to clarify the distinctive character of this take on the relation between philosophy and empirical science, let me briefly contrast it with the position recently advocated by Bennett and Hacker in their book *The Philosophical Foundations of Neuroscience*. According to their outlook, a philosophical investigation of consciousness differs in principle from an empirical one for which reason it is meaningless to suggest that the latter can challenge or even replace the former. Philosophy is not concerned with matters of fact, but with matters of meaning. The business of philosophy is with logical possibilities, not with empirical actualities. Its province is not the domain of empirical truth or falsehood, but the domain of sense and nonsense. To put it differently, philosophy clarifies what does and does not make sense. It investigates and describes the bounds of sense: that is, the limits of what can coherently be thought and said. The boundary between what does and what does not make sense, between what is meaningful and what transgresses the bounds of sense, is determined by the concepts we use, and the way philosophy can contribute to an investigation of the nature of the mind is consequently by clarifying our concept of mind and the way this concept is linked to related concepts (Bennett and Hacker 2003, 399, 402). The primary method of dissolving conceptual puzzlement is by carefully examining and describing the use of words – that is, we should investigate what competent speakers, using words correctly, do and do not say. Rather than engaging in first-order claims about the nature of things (which it can leave to various scientific disciplines) philosophy should consequently concern itself with the conceptual preconditions for any such empirical inquiries. Conceptual questions antecede matters of truth and falsehood. They are presupposed by any

---

[1] This is not to say, however, that Merleau-Ponty should be considered the ideal of how to integrate phenomenology and empirical science. In an article from 1996 Gallagher and Meltzoff pointed to certain shortcomings in Merleau-Ponty's use of developmental psychology, and back in 1979 the influential French phenomenological psychiatrist Tatossian criticized Merleau-Ponty for using empirical research in a speculative fashion. Tatossian wrote that if one really wants to grasp the phenomenological experience of the mentally ill person, one shouldn't remain in the ivory tower of the transcendental philosopher. Rather than just reading the specialized literature – as Merleau-Ponty had done – one should engage directly with the madman. This would, according to Tatossian, be the "genuine positivism" which Husserl was advocating (Tatossian 1979/1997, 12).

[2] A quite recent attempt to follow up on and further develop this specific take can be found in Evan Thompson's book *Mind in Life: Biology, Phenomenology and the Sciences of the Mind*.

scientific investigation, and any lack of clarity regarding the relevant concepts will be reflected in a corresponding lack of clarity in the questions posed, and hence in the design of the experiments intended to answer them (Bennett and Hacker 2003, 2). To put it more directly, empirical research that proceeds from conceptually flawed premises is likely to yield incoherent empirical questions and answers.[3] Bennett and Hacker then proceed by arguing that the relationship between conceptual and empirical issues is unidirectional, and that philosophy is of much greater importance to science than vice versa. For while philosophers can clarify the concepts used in science and thereby offer an immense service to science, it is a mistake to think that science could have much of an impact on philosophy. In fact, Bennett and Hacker even consider the supposition that scientific evidence may contravene a philosophical analysis ridiculous (Bennett and Hacker 2003, 404). In their view, we should not commit the mistake of confusing metaphysical or epistemological theories with empirical claims which can be corroborated by some *experimentum crucis*. Thus, the relation between philosophy and empirical science is a one-way enterprise. It is an application of ready-made concepts. There is no reciprocity, and there is no feedback. The application does not lead to a modification of the original analysis.

This is certainly one type of response to the challenge posed to philosophy by the revival of naturalism. But as I have just suggested, I don't think this is the way phenomenology has responded or should respond. To let an examination of ordinary language-use be our primary, if not exclusive, guide to a philosophical investigation of the mind is far too restrictive and underestimates the degree to which ordinary language reflects commonsense metaphysics. It blocks the way for concrete phenomenological analyses that might reveal aspects and dimensions of the mind that are not simply available to any reflection on common sense (consider for instance Husserl's investigations of the structures of time-consciousness or pictorial consciousness) and which at the same time might be challenged and enriched by descriptions provided by empirical science. By contrast, perhaps it is not entirely without reason that the

---

[3]Although I have reservations regarding Bennett and Hacker's depiction of the relation between philosophy of mind and empirical science, I find their criticism of certain grandiose tendencies in contemporary cognitive neuroscience quite to the point. The first two generations of modern neuroscientists, people such as Eccles and Penfield were neo-Cartesians. The third generation repudiated the dualism of their teachers and explicitly endorsed a form of physicalism. But as Bennett and Hacker point out, neuroscience has continued to remain bedevilled by a crypto-Cartesian and empiricist legacy. It might have replaced the immaterial Cartesian mind with the material brain, but it has maintained the dualism between brain and body, and thereby the logical structure of dualist psychology. Indeed, most of the neuroscientists who have castigated philosophy for its alleged failings – for not having accomplished anything scientifically worthwhile in its 2,500 year history – are unaware of the extent to which much of their own framework of thought has a questionable philosophical heritage. But as Bennett and Hacker argue, it is a simple fact, that the seventeenth-century philosophical conception of reality, of what is objective and what is subjective, of the nature of perception and its objects, has profoundly affected the ways in which brain scientists currently conceive of their own investigations (Bennett and Hacker 2003, 134). And as they then polemically ask, is what twenty-first century neuroscience can offer to philosophy simply a rehash of seventeenth-century epistemology and metaphysics (Bennett and Hacker 2003, 407)?

style of analytical philosophy defended by Bennett and Hacker has been accused of promoting a kind of semantic inertia and conceptual conservatism.

Empirical data can serve to challenge and corroborate theoretical analyses. Conversely, conceptual analysis can provide directions and tools for the empirical scientists and might also help in the design and development of experimental paradigms. But although empirical findings are important, we should obviously not overlook that they are open to interpretation. Their interpretation will usually depend upon the framework within which one is operating. Thus, the theoretical impact of an empirical case is not necessarily something that can be easily determined. Although I would argue that phenomenology should pay attention to empirical findings, this doesn't entail that it should also necessarily accept the (metaphysical and epistemological) interpretation that science gives of these findings.

It is important to encourage the exchange between phenomenology and empirical science, but the possibility of a fruitful cooperation between the two should not make us deny their difference. I see no incoherence in claiming that phenomenology should be informed by the best available scientific knowledge, while at the same time insisting that the ultimate concerns of phenomenology are transcendental philosophical and that transcendental philosophy differs from empirical science.

Philosophical Naturalism

So far, I have distinguished two different understandings of naturalization:

- The first radical proposal sees the naturalization of phenomenology as one that will eventually make phenomenology part of, or at least an extension of, natural science and it argues that this is something we should aim for. As I have made clear, I consider this suggestion to be misguided, since it de facto denies the legitimacy of methods and questions that are unique to philosophy. It wants to replace the transcendental clarification that phenomenology offers with an explanatory account. Were one to implement this strategy, one would by the same token abandon much of what makes phenomenology philosophically interesting. Phenomenology is basically, I would insist, a transcendental philosophical endeavor, and although one might ease the way for its naturalization by abandoning the transcendental dimension, one would not retain that which makes phenomenology a distinct philosophical discipline, strategy, and method.[4]
- The second more modest proposal argues that a naturalized phenomenology is the kind of phenomenology that engages in a meaningful and productive exchange with empirical science. Phenomenology can question and elucidate basic theoretical assumption made by empirical science, just as it might aid in the development

---

[4]Some might deny that post-Husserlian phenomenology is at all to be called transcendental, but I would disagree with this appraisal (cf. Zahavi 2008).

of new experimental paradigms. Empirical science can present phenomenology with concrete findings that it cannot simply ignore, but must be able to accommodate; evidence that might force it to refine or revise its own analyses.

At this point, however, it is necessary to briefly address an issue that so far has remained somewhat in the background. It could be argued that naturalism is far from being a univocal term, and that the real task facing phenomenology is to resist the attempt by metaphysical realists to monopolize the concept of nature. The real challenge would in other words be to rethink the very concept of nature – a concept about which Hume once declared "there is none more ambiguous and equivocal" (Hume 1888, 474).[5] In short, why make do with an unduly restricted conception of nature, why not recognize that there might be other kinds of naturalism than the one that takes it for granted that nature is exhausted by what natural science – as it is currently conceived – is capable of revealing to us (cf. McDowell 1996). To put it differently, maybe we should realize that it is mandatory to operate with a richer notion of nature, one that has room for such issues as meaning, context, perspectives, affordances and cultural sediments.

So far the point being made is simply that in discussing the relation between phenomenology and naturalism we shouldn't make the mistake of letting the concept of nature remain unexamined. In a parallel move, one might make a similar point regarding the notion of the transcendental.

One commentator has recently argued that Husserl through the 1920s and 1930s "became increasingly wide-reaching, even baroque, in his conception of the transcendental" (Moran 2002, 51). But rather than calling Husserl's notion of the transcendental baroque, perhaps it would be more to the point to realize that Husserl subjected the very notion of the transcendental to a far-reaching transformation. As I have argued elsewhere, Husserl's phenomenology is characterized by its attempt to modify the static opposition between the transcendental and the mundane, between the constituting and the constituted (Zahavi 2001, 2003). This was not an insight that Husserl only reached at the very end of his life. In a text which was written around 1914–15, and which has subsequently been published in *Husserliana* XXXVI, the volume entitled *Transzendentaler Idealismus. Texte aus dem Nachlass*, Husserl argued that actual being, or the being of actual reality, doesn't simply entail a relation to some formal cognizing subject, but that the constituting subject in question must necessarily be an embodied and embedded subject. Already in this period, Husserl was claiming that the subject in order to constitute the world must

---

[5]That there are speculative precedents to such a move in the history of philosophy is well known. It is no coincidence that Merleau-Ponty in the quote given above referred to Schelling. As the latter wrote in his *System of Transcendental Idealism*: "The highest consummation of natural science would be the complete spiritualizing of all natural laws into laws of intuition and thought. […] Nature's highest goal, to become wholly an object to herself, is achieved only through the last and highest order of reflection, which is none other than man; or, more generally, it is what we call reason, whereby nature first completely returns into herself, and by which it becomes apparent that nature is identical from the first with what we recognize in ourselves as the intelligent and the conscious" (Schelling 1978, 6).

necessarily be bodily embedded in the very world that it is seeking to constitute (Husserl 2003, 133). And as he then continued, the constitution of an objective world also requires that the subject stands in an essential relation to an open plurality of other embodied and embedded subjects (Husserl 2003, 135).

In *Les mots et les choses*, Foucault has argued that phenomenology exemplifies a type of modern discourse that in its investigation of experience seeks to both separate as well as integrate the empirical and the transcendental. It is an investigation of experience that in the face of positivism has tried to restore the lost dimension of the transcendental, but which at the same time has made experience concrete enough to include both body and culture. To Foucault it is quite clear that this modern type of transcendental reflection differs from the Kantian type by taking its point of departure in the paradox of human existence rather than in the existence of natural science. Although Husserl had apparently succeeded in unifying the Cartesian theme of the cogito with the transcendental motif of Kant, the truth is that Husserl was only able to accomplish this union in so far as he changed the very nature of transcendental analysis. When transcendental subjectivity is placed in the more fundamental dimension of time, the strict division between the transcendental and the empirical is compromised. The questions of validity and of genesis become entangled. It is this transformation that in Foucault's view has resulted in phenomenology's simultaneously promising and threatening proximity to empirical analyses of man (Foucault 1966, 331–336).

I think Foucault's diagnosis is correct (though I do not share his subsequent criticism of phenomenology). To put it differently, I think one must realize that the phenomenological notion of the transcendental differs from the Kantian one. This holds true not only for Husserl, but also for many of the post-Husserlian phenomenologists. As Merleau-Ponty would write in *Signs*:

"Now if the transcendental is intersubjectivity, how can the borders of the transcendental and the empirical help becoming indistinct? For along with the other person, all the other person sees of me – all my facticity – is reintegrated into subjectivity, or at least posited as an indispensable element of its definition. Thus the transcendental descends into history." (Merleau-Ponty 1964, 107)

The fact that phenomenology operates with an enlarged notion of the transcendental, the fact that it includes topics such as embodiment and intersubjectivity into its transcendental analysis, gives it an advantage in comparison with a more traditional Kantian type of transcendental philosophy. But, of course, it would also be fair to say that this enlargement and transformation generate new problems and challenges as well.

According to the proposal currently being considered a naturalization of phenomenology would entail a reexamination of the usual concept of naturalization and a revision of the classical dichotomy between the empirical and the transcendental. In short, according to the current proposal, a naturalization of phenomenology might not only entail a radical modification (rather than abandonment) of transcendental philosophy, but also a rethinking of the concept of nature – a rethinking that might ultimately lead to a transformation of natural science itself. Regardless of

how theoretically fascinating such a proposal might seem, it should, however, be obvious that the task is daunting and that there is still a long way to go.[6]

A more modest and considerable easier way to approach the issue regarding the relation between phenomenology and naturalism is to look concretely at how phenomenology and empirical science might engage in a fruitful and productive exchange. It is no coincidence that Husserl is reputed to have demanded small change (*Kleingeld*) from his students, and that he in a letter to Natorp wrote that he remained unsatisfied "as long as the large banknotes and bills are not turned into small change" (Husserl 1994, V. 56). Husserl stressed the importance of providing minute and careful analyses at the expense of developing ambitious and speculative systems. We should take this advice to heart. Of course, as in so many other cases, the proof of the pudding is in the eating, and any claims concerning a mutual enlightenment between phenomenology and empirical science must be demonstrated *in concreto*. However, this has already been done by an increasing number of people, and the present volume contains many further demonstrations.[7] In short, if you are genuinely interested in phenomenology and the problem of naturalization I recommend that you stop reading this article and instead turn to some of the other contributions in the volume.[8]

References

Aikin SF (2006) Pragmatism, naturalism, and phenomenology. Hum Stud 29:317–340
Bennett MR, Hacker PMS (2003) Philosophical foundations of neuroscience. Blackwell, Oxford
Crick F (1995) The astonishing hypothesis. Touchstone, London
Foucault M (1966) Les mot et les choses. Gallimard, Paris
Gallagher S (1997) Mutual enlightenment: recent phenomenology in cognitive science. J Conscious Stud 4(3):195–214
Gallagher S (2003) Phenomenology and experimental design. J Conscious Stud 10(9–10):85–99

---

[6]It should be noted though that there are those who argue that the full theoretical implications of Einstein's and Bohr's revolutions in physics are yet to be realized. In 1922, Moritz Schlick gave a talk where he argued that the general theory of relativity had disconfirmed transcendental philosophy and vindicated empiricist philosophy. This view has found much resonance, but as Ryckman has argued in his recent book *The Reign of Relativity: Philosophy in Physics 1915–1925* it happens to be quite incorrect. As Ryckman points out, the outstanding mathematician Herman Weyl who was one of Einstein's colleagues in Zürich, and who contributed decisively to the interpretation and further development of both the general theory of relativity and the field of quantum mechanics, did not only draw quite extensively on Husserl's criticism of naturalism, but was also deeply influenced by Husserl's transcendental idealism. Thus, it is by no means obvious that some of the decisive developments in theoretical physics really leave our standard conception of subjectivity, objectivity and knowledge untouched.

[7]For some of my own contributions, see Zahavi (2005) and Gallagher and Zahavi (2008).

Gallagher S, Meltzoff A (1996) The earliest sense of self and others: Merleau-Ponty and recent developmental studies. Philos Psychol 9:213–236

Gallagher S, Zahavi D (2008) The phenomenological mind. An introduction to philosophy of mind and cognitive science. Routledge, London

Heidegger M (1978) Wegmarken. Vittorio Klostermann, Frankfurt am Main

Hume D (1888) A treatise of human nature. Clarendon, Oxford

Husserl E (1970) The crisis of European sciences and transcendental phenomenology. An introduction to phenomenology (trans: Carr D). Northwestern University Press, Evanston, IL

Husserl E (1977) Phenomenological psychology: lectures, summer semester, 1925 (trans: Scanlon J). Martinus Nijhoff, The Hague

Husserl E (1982) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. First book. General introduction to a pure phenomenology (trans: Kersten F). Martinus Nijhoff, The Hague

Husserl E (1984) Einleitung in die Logik und Erkenntnistheorie. Vorlesungen 1906/07. Husserliana XXIV. Martinus Nijhoff, Dordrecht

Husserl E (1987) Aufsätze und Vorträge (1911–1921). Husserliana XXV. Martinus Nijhoff, Dordrecht

Husserl E (1989) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. Second book. Studies in the phenomenology of constitution (trans: Rojcewicz R, Schuwer A). Kluwer, Dordrecht

Husserl E (1994) Briefwechsel, Husserliana Dokumente III/1–10. Kluwer, Dordrecht

Husserl E (1999) Cartesian meditations: an introduction to phenomenology (trans: Cairns D). Martinus Nijhoff, The Hague

Husserl E (2001) Logical investigations I-II (trans: Findlay JN). Routledge, London

Husserl E (2003) Transzendentaler Idealismus. Texte aus dem Nachlass (1908–1921). Husserliana XXXVI. Kluwer, Dordrecht

Lewis M (2003) The development of self-consciousness. In: Roessler J, Eilan N (eds) Agency and self-awareness. Oxford University Press, Oxford, pp 275–295

Lutz A (2002) Toward a neurophenomenology as an account of generative passages: a first empirical case study. Phenomenol Cogn Sci 1:133–167

Lutz A, Thompson E (2003) Neurophenomenology: integrating subjective experience and brain dynamics in the neuroscience of consciousness. J Conscious Stud 10:31–52

McDowell J (1996) Two sorts of naturalism. In: Hursthouse R, Lawrence G, Quinn W (eds) Virtues and reasons: Philippa Foot and moral theory. Clarendon, Oxford, pp 149–179

Merleau-Ponty M (1962) Phenomenology of perception (trans: Smith C). Routledge & Kegan Paul, London

Merleau-Ponty M (1963) The structure of behavior (trans: Fisher AL). Duquesne University Press, Pittsburgh

Merleau-Ponty M (1964) Signs (trans: McCleary RC). Northwestern University Press, Evanston, IL

Merleau-Ponty M (2000) Nature: course notes from the Collège de France. Northwestern University Press, Evanston, IL

Moran D (2002) Making sense: Husserl's phenomenology as transcendental idealism. In: Malpas J (ed) From Kant to Davidson: philosophy and the idea of the transcendental. Routledge, London, pp 48–74

Roy J-M, Petitot J, Pachoud B, Varela FJ (1999) Beyond the gap: an introduction to naturalizing phenomenology. In: Petitot J, Varlea FJ, Pachoud B, Roy J-M (eds) Naturalizing phenomenology. Stanford University Press, Stanford, CA, pp 1–83

Ryckman T (2005) The reign of relativity: philosophy in physics 1915–1925. Oxford University Press, New York

Schelling FWJ (1978) System of transcendental idealism (1800) (trans: Heath P). University of Virginia Press, Charlottesville, VA

Sellars W (1963) Science, perception and reality. Routledge and Kegan Paul, London

Tatossian A (1979/1997) La phénoménologie des psychoses. L'art du comprendre, Paris

Thompson E (2007) Mind in life: biology, phenomenology, and the sciences of mind. Harvard University Press, Cambridge, MA

Varela F (1996) Neurophenomenology: a methodological remedy to the hard problem. J Conscious Stud 3:330–350

Varela F (1997) The naturalization of phenomenology as the transcendence of nature. Searching for generative mutual constraints. Alter 5:355–381

Williams B (2005) Descartes: the project of pure enquiry. Routledge, London

Zahavi D (2001) Husserl and transcendental intersubjectivity. A response to the linguistic-pragmatic critique. Ohio University Press, Athens

Zahavi D (2003) Husserl's phenomenology. Stanford University Press, Stanford, CA

Zahavi D (2005) Subjectivity and selfhood: investigating the first-person perspective. MIT Press, Cambridge, MA

Zahavi D (2008) The phenomenological tradition. In: Moran D (ed) Routledge companion to twentieth-century philosophy. Routledge, London, pp 661–692

Phenomenology and Non-reductionist Cognitive Science

Shaun Gallagher

> Psychology was defined by William James as 'the science of mental life', but psychologists seem to have developed a great reluctance to study 'mental life' by any but the most indirect routes. The most extreme example of this lack of directness is seen in functional brain imaging. For some researchers it seems that, if you have a brain scanner, then you no longer need to study mental activity because brain activity is a truly objective alternative. (Frith 2002, 374)

The basic argument of this chapter, and more generally of this volume, is that it is possible to have a non-reductionist science of the embodied mind that is superior in many ways to any reductionist science that uses only "indirect routes" to, and often fails to arrive at experience. More specifically, in recent years, arguments have been put forward that we can pursue this non-reductionist science to the extent that phenomenology, or alternative introspective methods that can provide access to a methodologically controlled description of first-person experience, can be employed in experimental science (Frith 2002; Gallagher 2002, 2003; Gallagher and Overgaard 2005; Gallagher and Sørensen 2006; Jack and Roepstorff 2002; Schooler 2002; Varela and Shear 1999; also see especially the papers collected in the two-volume *Trusting the Subject*, Jack and Roepstorff 2003; Roepstorff and Jack 2004). Despite this growing but cautious agreement about the importance of first-person approaches, there are still questions about precisely what these methods are and how they are to be used. There are also doubts and objections, most famously summarized by Dennett (2001): "First-person science of consciousness is a discipline with no methods, no data, no results, no future, no promise. It will remain a fantasy." For purposes of this chapter I set aside such objections (see Noë 2007 for ongoing debates), and focus on the varieties of first-person approaches that can contribute to cognitive science.

S. Gallagher
Philosophy and Cognitive Science, Institute of Simulation and Training,
University of Central Florida (USA) and Philosophy Department
University of Hertfordshire (UK)

Introspection and Beyond

Price and Aydede (2005), for example, argue that introspection is indispensible and that it is already extensively used as a part of third-person studies. But what is the nature of the introspection that is already practiced in science? Schooler (2002), in his account, lists various terms used to signify introspection. In most cases, what is indicated is a 'second-order' reflective access relative to first-order phenomenal experience. This kind of second-order reflective activity can be very simple, and as such can be found in a large variety of experiments actually designed to minimize dependency on introspection. In such cases, experimenters might ask their subjects for quick reports about what they experience. '*Do you experience* (see, hear, feel, etc.) *X or not*'? In some cases, to avoid the effects of verbal misinterpretation the subject is asked to simply push a button once she experiences *X*. This still depends on a quick and minimal introspection of the first-order experience (seeing, hearing, feeling, etc.) to be reported. If one instructs a subject to push a button, or say 'now' when they see the light come on, then the subject is reporting about the light, but also about their visual experience. Even if one instructs the subject in a way that carefully avoids mention of an experiential state: "Push the button when the light comes on," the only access that the subject has to the fact of the light coming on is by way of her experience of the light coming on. In this sense the first-person perspective in inherent in experiments that depend on subjective reports.

More generally, however, should we say that the first-person reports of a subject are necessarily introspective? For example, I may ask the subject to say 'now' when they see the light come on. How precisely does the subject know when they see the light come on? Do they reflectively introspect their experience looking for the particular visual state of seeing the light come on? Or do they simply see the light come on and report that? Is it possible that we can report on what we experience without employing introspection?[1]

There is a long tradition in philosophical phenomenology (specifically the tradition that follows Husserl)[2] that answers in the affirmative. We can report on what we experience without using introspection because we have an implicit, non-introspective, first-person, pre-reflective self-awareness of our own experience. At the same time that I see the light, I know that I see the light. This knowledge of seeing the light is not based on reflectively or introspectively turning our attention to our own experience. It is rather built into our experience as an essential part of it, and it is precisely that which defines our experience as conscious experience. On this view, I consciously experience the light coming on just as I see the light coming on. I don't have to verify through introspection that I have just seen the light come on,

---

[1]Marcel (1998) has demonstrated good reason to be cautious even about this kind of minimal procedure. Specifically, across different report modes (button push, eye blink, verbal '*Yes*') it is possible for different reports to be generated for the very same trial.

[2]As is the practice throughout this volume, reference to phenomenology means the approach originating with Husserl and developed in the work of his followers.

since my first-order phenomenal experience is already something of which I am aware in the very experience of it.

Introspective self-observation of a more reflective variety may be called for in some experiments. For instance, if asked to report on an emotion, a subject is required to make considered judgments about her own first-order phenomenal experience. In other cases, when the experimenter asks what exactly the subject does experience, rather than asking whether or not the subject experiences a specific *X*, the subject may use reflective introspection which involves an interpretation of the experience. One important question here concerns the basis for that interpretation. If no instruction is given, a naïve subject is likely to give their report in categories taken from folk psychology. To avoid the subjective aspects of such interpretations, the scientist may decide to instruct the subject, or to provide a pre-established set of categories from which the subject chooses the relevant interpretation. This simply leads to a further question. What is the source of the pre-established categories? It seems quite possible that even if their source consists in previous scientific work, they ultimately can be traced back to folk psychological roots (Gallagher 1997). So even in the attempt to be objective and to avoid what seems to be a possible subjective bias in introspection, there may still be a significant bias if the categories in play are not derived by means of a controlled method. Introspective self-observation must also answer to the possibilities of temporal and translation dissociations noted by Schooler (2002) and many of the traditional critics of introspection.

There are two good responses to such issues, both of which involve phenomenology.[3] The first takes a more systematic approach to introspection by using controlled procedures that allow the experimental subject's first-person experience to inform the experimental analysis. This approach is exemplified by neurophenomenology. By instructing subjects to set aside standard (folk psychological) conceptions and theories, instructing them to focus on the first-order experience itself, and asking 'open' questions, neurophenomenology, employing the methods of systematic phenomenology, allows subjects themselves to define the proper analytic categories. In other words, categories are generated in the very first-order experience that the experimenter wants to know about, rather than in some other, often anonymous, first-order experience, the understanding of which, and relevance of which are a matter of interpretation. Neurophenomenology also employs triangulation, a strategy recommended by Jack and Roepstroff (2002). That is, it combines qualitative phenomenological methods with quantitative measurements of brain processes and a dynamical interpretation of the data.

This kind of procedure will not work in every type of experiment, however, nor will it necessarily deliver the best results even in those where it is possible. As Marcel (1998) notes, for example, in attempts to measure the effect of unconscious processing, as in priming, tasks that use phenomenological procedures, no matter how well they are defined procedurally, might not reveal the degree of priming effect as clearly as procedures that measure the effect indirectly. Also we can mention

---

[3]For more on the distinction and relation between introspection and phenomenology, see Gallagher and Zahavi (2008); Overgaard, Ramsøy and Gallagher (2008); and M. Overgaard (this volume).

cases that may involve certain pathologies where training subjects in phenomenological method may be impossible. A second phenomenological approach that does not involve training experimental subjects in phenomenological method may thus have wider application. This approach involves "front-loading" phenomenological insights into experimental design (Gallagher 2003).

Neurophenomenology

Neurophenomenology is based on an approach first outlined by Francisco Varela (1996). Following this approach, phenomenological methods have been productively incorporated into the protocol of experiments on perception (Lutz et al. 2002), and on epilepsy by the Le Van Quyen–Varela group (Le Van Quyen et al. 1997a, b; 1999, 2001a, b; Le Van Quyen and Petitmengin 2002; see Petitmengin, this volume). Such procedures hold great promise for specific types of experiments, including those that use neuroimaging.

Neurophenomenology attempts to integrate three elements: (1) phenomenological analysis of experience; (2) dynamical systems theory; and (3) empirical experimentation on biological systems. Varela (1996) maintains that for purposes of neurophenomenology, both empirical scientists and experimental subjects ought to receive some level of training in phenomenological method (also see Roy et al. 1999). This training includes learning to practice the phenomenological *epoche*, that is, the setting aside or "bracketing" of opinions or theories that a subject may have about experience or consciousness. It involves shifting our attention from *what* we experience to *how* we experience. Lutz et al. (2002) have shown the practicality of this approach with some success.

Lutz, Varela, and their colleagues studied the highly variable successive brain responses to repeated and identical stimulations found in many empirical testing situations that target specified cognitive tasks. Their hypothesis was that this variability is generated in mental fluctuations due to the subject's attentive state, spontaneous thought-process, strategy decisions for carrying out the task, etc. These *subjective parameters* (SPs) include distractions, cognitive interference, etc. To control for SPs is difficult and they are usually classified as unintelligible noise (Engel et al. 2001), ignored, or neutralized by a method of averaging results across a series of trials and across subjects. Lutz and his colleagues decided to take SPs more seriously. They combined a process of trained phenomenological reflection with the dynamical analysis of neural processes measured by EEG in a paradigm involving a 3D perceptual illusion. Random-dot patterns with binocular disparities (autostereogram) were presented on computer screen. By visually manipulating these dots, subjects were able to see a 3-D illusory geometric shape emerge (depth illusion). They were instructed to press a button with their right hand as soon as the shape had completely emerged. After the button push, the subjects gave a brief verbal report of their experience.

Lutz and his colleagues used the first-person data not simply as more data for analysis, but as contributing to an organizing analytic principle. Specifically, the trials

were clustered according to first-person descriptive reports concerning the experience of SPs, and for each cluster separate dynamical analyses of electrical brain activity, recorded by EEG, were conducted. The results were different and significant in comparison to a procedure of averaging across trials.

The phenomenological part of the experiment involved the development of descriptions (refined verbal reports) of the SPs through a series of preliminary or practice trials. In the preliminary training subjects became knowledgeable about their own experience, defined their own categories descriptive of the SPs, and could report on the presence or absence or degree of distractions, inattentive moments, cognitive strategies, etc. Based on the subject's own trained reports, descriptive categories were defined *a posteriori* and used to divide the trials into phenomenologically based clusters. For example, with regard to the subject's experienced readiness for the stimulus, the results specified three readiness states:

- **Steady readiness (SR):** subjects reported that they were 'ready', 'present', 'here', 'well-prepared' when the image appeared on the screen and that they responded 'immediately' and 'decidedly'.
- **Fragmented readiness (FR):** subjects reported that they had made a voluntary effort to be ready, but were prepared either less 'sharply' (due to a momentary 'tiredness') or less 'focally' (due to small 'distractions', 'inner speech' or 'discursive thoughts').
- **Unreadiness (SU):** subjects reported that they were unprepared and that they saw the 3D image only because their eyes were correctly positioned. They were surprised by it and reported that they were 'interrupted' by the image in the middle of an unrelated thought.

Subjects used these categories during the main trials when the experimenters recorded both electrical brain activity and the subject's own report of each trial. The reports during the main trials revealed subtle changes in the subject's experience due to the presence of specific SPs, reflecting the subject's attention level.

The clustered first-person data were then correlated with both behavioral measures (reaction times) and dynamic descriptions of the EEG-recorded transient patterns of local and long-distance synchrony occurring between oscillating neural populations, specified as a dynamic neural signature (DNS). Such coherent temporal patterns occur during ongoing activity related to top-down factors such as attention, vigilance, or expectation. Lutz et al. were able to show that distinct SPs, described in the subjects' trained phenomenological reports, correlated with distinct DNSs just prior to presentation of the stimulus.

On the neurophenomenological model, subjects are trained to deliver consistent and clear reports of their experience, and to attain intuitions of the descriptive structural invariants of an experience. They are asked to bracket their ordinary attitudes in order to shift their attention from *what* they experience to *how* they experience it. Subjects are asked to suspend their beliefs, common-sense opinions, or theories about experience. If, for example, subjects are well trained in neuroscience or in psychoanalysis, they are asked to put that knowledge aside, to bracket it, and to attend directly to their experiences rather than to interpretations of their experiences. They are then asked to practice and to become acquainted with the kind

of experiences that they will be tested on. In the end, their descriptions are subject to intersubjective verification.

These phenomenological procedures can be either self-induced by subjects familiar with the methods, or guided by the experimenter through open questions – questions not directed at opinions or theories, but at experience (see Vermersch 1994; Petitmengin-Peugeot 1999; this volume).

> To train the subjects, open questions were asked to try to redirect their attention towards their own immediate mental processes before the recordings were taken. … For example: Experimenter: 'What did you feel before and after the image appeared?' Subject S1: 'I had a growing sense of expectation, but not for a specific object; however when the figure appeared, I had a feeling of confirmation, no surprise at all'; or subject S4: 'it was as if the image appeared in the periphery of my attention, but then my attention was suddenly swallowed up by the shape'. (Lutz et al. 2002)

Open questions posed immediately after the task assist the subject in a reflective assessment of the degree of attention implemented during the task. Subjects are re-exposed to the stimuli until they find "their own stable experiential invariants" which they use to describe the specific elements of their experiences, in this case, the SPs. These invariants then become the defining elements of intersubjectively agreed upon phenomenological clusters that are used as analytic tools in the main trials, and correlated with objective measurements of behavior (reaction times) and brain activity (see Lutz 2002, for further analysis).

Front-Loading Phenomenology

Merleau-Ponty frequently used phenomenological insights to reinterpret experimental results. In such cases, phenomenology can take on a critical function, offering correctives to various theoretical interpretations of the empirical data. Although this kind of after-the-fact phenomenological reinterpretation can be theoretically productive, in that it develops alternative interpretations, unless these interpretations are subject to further empirical testing, they remain unverified. This simply brings us back to the question of how to incorporate such phenomenological insights into experimental studies.

Another problem with the idea of using phenomenology to reinterpret experimental results can be seen in regard to pathological case studies. Although Merleau-Ponty (1962) offers a brilliantly conceived reinterpretation of the case of Kurt Goldstein's brain-damaged patient Schneider, there was never any verification by way of new testing of Schneider.[4] As a result, Merleau-Ponty's account of the case remains simply one of several possible theoretical accounts. In a very practical way this suggests the inadequacy of this kind of retrospective reinterpretation if phenomenologists are not working directly with and along side psychologists and neuroscientists in

---

[4] Schneider's brain damage was extremely complex, and was studied by Goldstein between 1915 and 1930.

the experimental context. And once again this brings us back to the question of how specifically to incorporate phenomenology into the experimental context.

This is a question addressed by what I have called "front-loading" phenomenology (Gallagher 2003; Gallagher and Sørensen 2006; Gallagher and Zahavi 2008). Rather than starting with the empirical results (as one would do in retrospective reinterpretation), or with the training of subjects (as one would do on the neurophenomenological approach) front-loading phenomenology starts with the experimental design. Phenomenological insights (concepts, distinctions) developed in separate phenomenological analyses, or in neurophenomenological experiments, are used to inform the design of experiments. To front load phenomenology, however, does not mean to simply presuppose phenomenological results obtained by others. Rather it involves testing those results and more generally a dialectical exchange between phenomenological insights and preliminary trials that will specify or extend these insights for purposes of the particular experiment or empirical investigation. It is a natural consequence of front-loading phenomenology that, as in the neurophenomenological model, the phenomenology becomes part of the analytic framework for interpreting the results, and not just part of the data to be analyzed.

Some examples will help to clarify this approach. The following experiments involve brain imaging, but, in order to eliminate any confusion about where precisely the phenomenological contribution lies, the first two do not involve first-person or introspective reports.[5] In regard to each experiment, the phenomenological analysis is done prior to the experiments, and the results of that analysis are used to work out the experimental design.

Three recent experiments (Farrer and Frith 2002; Chaminade and Decety 2002; Farrer et al. 2003) set out to identify the neural correlates of the sense of agency. The starting point for each was a phenomenological distinction between the sense of agency (SA) and the sense of ownership (SO) (Gallagher 2000a). In part, the experimental design required that the experimenters control for SO as they attempted to study SA.

In the normal experience of intentional action the sense of ownership for the action and the sense of agency for the action are close to indistinguishable. In the case of involuntary action, however, they are easily distinguished. If, for example, someone pushes me from behind, I sense that it is my body that is moving – it is *my* movement and I experience ownership for the movement – but I do not experience self-agency for the movement (I have no sense that I intended or caused the movement). The phenomenology, however, is more complicated, because we need to distinguish between the first-order phenomenal level of experience of agency (or ownership), and a higher-order cognitive judgment or attribution of

---

[5] There are experiments that rely on front-loading phenomenology that do employ introspective reports. Brosted Sørensen (2005), for example, following Nielsen (1963), designed an alien-hand experiment to test visual *versus* proprioceptive awareness of the body in bulimic patients, as evidenced by first-person reports. For the experimental design he relies on phenomenological distinctions between body image and body schema (see Gallagher and Cole 1995; Paillard 1997, 1999; Gallagher 2005), and between sense of agency and sense of ownership (Gallagher 2000a, b).

agency (or ownership). Graham and Stephens (1994; see Stephens and Graham 2000), for example, distinguish between "attributions of subjectivity" (ownership) and "attributions of agency" on the level of higher-order, reflective or introspective attribution. I can attribute an action to myself based on a reflective awareness of what I have done, for example, in the case when someone asks me about it. One could think of this as involving a level of narrative where I can give an account of my actions, and perhaps provide reasons for acting in the way that I have done. The distinction between SA and SO, however, is more basic and applies to first-order phenomenal experience. That is, in the case of involuntary movement, I directly experience the movement as happening to me (SO), but not as caused by me (no SA). SO and SA are seemingly (and in the case of phenomenal experience, 'seemingly' means 'really') built into experience. They are part of a pre-reflective (non-conceptual) self-awareness implicit to the experience of action. Indeed, this is usually the basis for any higher-order attributions of subjectivity and agency.

The goal of the experiments was to determine what neuronal processes generate the first-order phenomenal experience of SA. Given that SA is implicit in first-order phenomenal experience rather than the product of second-order introspective attribution, then the experimenters were looking for a more basic set of primary processes that are activated in motor control mechanisms rather than in areas that may be responsible for higher-order cognitive processes. Insofar as the phenomenological distinction between SO and SA informed the design of these experiments, then, they are good examples of front-loaded phenomenology. There is more to it than that, however. These experiments also show (1) how difficult it can be to keep the phenomenological distinction straight; and (2) how the experimental science can also assist phenomenology in drawing out the complexity of such phenomenological distinctions.

The original context for making the distinction between SA and SO related to studies of schizophrenia, where in positive symptoms of delusions of control the subject seemingly has trouble with SA, but retains SO for the actions (Gallagher 2000a, b). In this precise context, SA was defined in terms of motor-control, and specifically in terms of whether I have control over my bodily movements. The distinction between SA and SO, then, was cast in precisely this way. SO was associated with sensory feedback (e.g., proprioception) that results from bodily movement and that is present even in involuntary movements and in the experiences of schizophrenic delusions of control. SA was associated with pre-action efferent control processes (e.g., motor commands, efference copy) activated in voluntary action. This way of defining the difference is a relatively narrow one, focused just on motor control issues. SA, however, is more complex than this (Gallagher 2008; in press). It also involves, at the very least, a sense of accomplishing some effect in the world. My experience of my action is not purely a motor experience – it is also world-involving, since it is directed at some task in the world. By my action I am trying to accomplish something, and if I accomplish the task, for example, SA may be reinforced and therefore a stronger experience. It is clearly possible to have a sense of agency informed primarily by my sense of accomplishing what I set out to do. Indeed, generally our awareness of action is not focused on our bodily movements,

but on the intentional task at hand. This complexity has to be taken into account in the phenomenology, and if it is not it can easily lead to a confusion about what precisely SA is. This is what happens in the experiments.

Chaminade and Decety (2002)

In the positron-emission tomographic study by Chaminade and Decety (2002), subjects were asked to move a joystick to control an icon on a computer screen to accomplish several tasks, the critical ones being task A and task B.

Task A: Subject moved the icon and observed another subject's icon following (leader task)

Task B: Subject followed another subject's icon with her own icon (follower task)

The subjects either led (A) or followed (B) the other's icon, "in a computerized environment free of explicit reference to body parts…. The sense of ownership, related to motor control, and the sense of agency, related to the intentional aspect, can be segregated in the analysis" (p. 1977). The reasoning here is that movement was required in each case and SO would therefore be constant. That is, the subject experiences her own movement in each case, but, Chaminade and Decety propose, SA would be different for task A versus task B. It soon becomes clear that they emphasize "a computerized environment free of explicit reference to body parts" because they understand SA to be linked to the intentional aspect of action rather than to bodily movement. On their view, SA varies across the tasks as the intention changes across the tasks, while SO remains constant because that is tied to afferent feedback from the consistent bodily movement.

One questionable assumption here is that A (leading) would generate SA, while B (following) would not. One obvious objection to this is that in both tasks (A and B) the subject may have SA for the intentional aspect of accomplishing the task. If asked to report (in fact, subjects were not asked to report), they might have said: "My task in A is to lead, and I have done so; and my task in B is to follow, and I have done so. I felt myself to be the agent of both actions." Accordingly, differential activation of the identified brain areas may be for something other than agency. Furthermore, even if we ignore the intentional aspect and focus on the motor aspect, the subject will experience SA in both cases because in both A and B the subject moves his hand to control the joystick (see Tsakiris and Haggard 2005).

Despite these complications, the researchers found activation in the pre-supplemental motor area (SMA) and the right inferior parietal cortex (IPC) in task A. They suggested that these areas are responsible for the sense of self-agency. In contrast, activation of the left IPC and the right precentral gyrus indicated a lack of self-agency or the attribution of agency to the other. They also noted the relevance to schizophrenia – "abnormal increased activity in the right inferior parietal cortex has been observed in schizophrenic patients experiencing passivity phenomenon" (p. 1978).

Even if this study begins with a phenomenological distinction between SA and SO, it soon becomes clear that further phenomenological work needs to be done to

sort out the difference between SA for the intentional aspect (the task, what gets done in the world, or on the computer screen) and SA for motor control. Further phenomenological clarification, then, would lead to newly designed experiments that may get us closer to understanding the brain processes responsible for SA.

Farrer and Frith (2002)

Chlöe Farrer and Chris Frith (2002) followed a similar logic as Chaminade and Decety in their fMRI experiment: "Subjects manipulated a joystick [to drive a colored circle moving on a screen to specific locations on the screen]. Sometimes the subject caused this movement and sometimes the experimenter. This paradigm allowed us to study the sense of agency without any confounding from the sense of ownership. To achieve this, subjects were requested to execute an action during all the different experimental conditions. By doing so the effect related to the sense of ownership (I am performing an action) would be present in all conditions and would be canceled in the various contrasts" (p. 597). In this case, as in the previous experiment, SA is thought to vary because the intentional aspect of action varies. So again, and for the same reasons, the claim is that SO remains constant while SA varies. But again, one could object that since in each case the subject is required to move the joystick, SA (defined in terms of motor control) for that action must result.

In some trials the subject is informed before each task whether the movement will be his or not his. He is asked to perform the task regardless of whether what happens on the screen is known to the subject to be his action or the action of someone else. In the case where he is not the agent for the task represented on the computer screen, the right IPC is activated. When he does know that he is causing the action on the screen, his anterior insula is activated bilaterally. Thus, the experimenters identify activation in the anterior insula as the correlate of SA.

Despite the fact that Farrer and Frith link SA to the intentional aspect of the task, when it comes to providing a theoretical explanation about why the anterior insula should be involved in generating SA they revert to a explanation more consistent with linking SA to motor control – that is, they explain the involvement of the anterior insula in terms of motor control.

> Why should the parietal lobe have a special role in attributing actions to others while the anterior insula is concerned with attributing actions to the self? The sense of agency (i.e., being aware of causing an action) occurs in the context of a body moving in time and space…. There is evidence that both the inferior parietal lobe and the anterior insula are representations of the body … the anterior insula, in interaction with limbic structures, is also involved in the representation of body schema…. One aspect of the experience of agency that we feel when we move our bodies through space is the close correspondence between many different sensory signals. In particular there will be a correspondence between three kinds of signal: somatosensory signals directly consequent upon our movements, visual and auditory signals that may result indirectly from our movements, and last, the corollary discharge associated with motor commands that generated the movements. A close correspondence between all these signals helps to give us a sense of agency. (pp. 601–602)

If this is the case, however, the fact that for each task the subject is required to move does complicate things. As for the previous experiment, further phenomenological clarification of SA could provide a better design for these experiments.

Farrer et al. *(2003)*

In a follow-up experiment Farrer et al. actually asked for the subject's report. Moreover, in this study, all questions about agency were focused on bodily movement. Subjects were not given any intentional task to carry out other than making random movements using a joystick, and the focus of their attention was directed toward a virtual (computer image) hand that either did or did not represent their own hand movements, although at varying degrees of rotation relative to true position of the subject's hand. In other words, they moved their own hand but saw a virtual hand projected on screen at veridical or nonveridical angles to their own hand; the virtual hand was either under their control, or not. Subjects were then asked about their sense of agency for control of the virtual hand movements.

The less the subject felt in control, the higher the level of activation in the right IPC (consistent with Farrer and Frith 2002). The more the subject felt in control, the higher the level of activation in the right *posterior* insula (in contrast with Farrer and Frith [2002], where SA was associated with activation of the right *anterior* insula). Noting this difference, Farrer et al. state: "We have no explanation as to why the localization of the activated areas differ in these studies, except that we know that these two regions are densely and reciprocally connected" (2003, p. 331). Clearly one possible explanation is that the shift of focus from accomplishing a computer screen task (in Farrer and Frith) to controlling bodily movement (in Farrer et al.) might change the phenomenon that is being studied.

If, in this case, we understand SA as generated by bodily movement or motor control, rather than the intentional aspect of action, since bodily movement seems to be the only thing at stake in this experiment, then the fact that the subject moves his own hand in all instances suggests that there should be no clear variation in SA, and no way to discriminate SO from SA. Indeed, one might think that the difference in visual feedback based on varying degrees of rotation, might have more to do with SO. Again, complicating the issue, when it comes to a theoretical explanation of why the IPC is involved in the question of agency, the authors cite evidence that pertains to ownership rather than agency: "Lesions of the inferior parietal cortex, especially on the right side, have been associated with delusions about the patient's limb that may be perceived as an alien object or as belonging to another person" (Farrer et al. 2003, p. 329). Such delusions are clearly about SO rather than SA.

These three experiments are examples that involve front-loading phenomenological distinctions (between SA and SO) into the experimental design. These experiments also make it clear that what we learn from the experiments may also be instructive for the phenomenologist. Clearly, more refined phenomenological

considerations about the sense of agency and sense of ownership, about the relation between bodily movement and intentional aspect, can further inform and improve future experiments on these phenomena. Generally, we should not think of experiments as simply accepting the phenomenological description. Rather they can test and verify that description, and either extend its application, or send it back to the phenomenological drawing board. In such cases we would have a mutual enlightenment between phenomenology and the empirical science.

Conclusion

A front-loading phenomenological approach does its primary work in clarifying the phenomenological distinctions and insights that might contribute to experimental design and interpretation. The neurophenomenological approach integrates phenomenological methodology into the actual workings of the experiment and offers correlations with behavioral and neuroscientific data. The idea is not, I think, that every experiment has to privilege the first-person data internal to the experiment, as long as the significance of first-person experience gets taken into account at some point in the process. In the case of front-loading phenomenology, the first-person data is taken seriously in the phenomenological analysis that serves to set up the design. In all cases, it is good scientific practice to understand and to introduce controls on the various experiential categories that may be involved, both in experiential reports and in the interpretation of those reports. It is not a matter of blindly trusting the subject, or blindly distrusting the subject.

This kind of scientific use of phenomenology need not challenge the original scope of phenomenology. Although Husserl had defined phenomenology as a transcendental (non-naturalistic) enterprise, the idea that science should ignore the results of his transcendental analyses would be entirely inconsistent with his own intent. He suggested, quite clearly, that "every analysis or theory of transcendental phenomenology – including … the theory of the transcendental constitution of an objective world – can be developed in the natural realm, by giving up the transcendental attitude" (1970, p. §57). This would be an extension of phenomenology rather than a rejection of its original idea. But by using phenomenology in experimental science, the gains do not go to science alone. Rather, empirical science can test and verify phenomenological descriptions and can extend its application.

If there is something like subjective experience – and can anyone really deny this? – then good science should not ignore it. The use of phenomenology in the empirical cognitive sciences reinforces the importance of first-person experience and thereby undermines the reductionist tendencies that one often finds in scientific theory. Since we are more than just a bunch of neurons, one requires good methods that will allow us to sort out what the "*more*" is. Phenomenologically enlightened ways of understanding first-person experience and how it affects experimental results just is one such method.

References

Chaminade T, Decety J (2002) Leader or follower? Involvement of the inferior parietal lobule in agency. NeuroReport 13(1528):1975–1978

Dennett D (2001) The fantasy of first-person science. http://ase.tufts.edu/cogstud/papers/chalmersdeb3dft.htm accessed on 25 August, 2009

Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci 10:704–716

Farrer C, Frith CD (2002) Experiencing oneself vs. another person as being the cause of an action: the neural correlates of the experience of agency. NeuroImage 15:596–603

Farrer C, Franck N, Georgieff N, Frith CD, Decety J, Jeannerod M (2003) Modulating the experience of agency: a positron emission tomography study. NeuroImage 18:324–333

Frith CD (2002) How can we share experiences? Trends Cogn Sci 6(9):374

Gallagher S (in press) Multiple aspects of agency. New Ideas Psychol

Gallagher S (2008) Agency, free will, and psychopathology. In: Parnas J, Kenneth S. Kendler (eds) Philosophical issues in psychiatry: natural kinds, mental taxonomy and the nature of reality. Johns Hopkins University Press, Baltimore, MD, pp 286–312

Gallagher S (2005) How the body shapes the mind. Oxford University Press, Oxford

Gallagher S (2003) Phenomenology and experimental design. J Conscious Stud 10(9–10):85–99

Gallagher S (2002) Experimenting with introspection. Trends Cogn Sci 6(9):374–375

Gallagher S (2000a) Philosophical conceptions of the self: implications for cognitive science. Trends Cogn Sci 4(1):14–21

Gallagher S (2000b) Self-reference and schizophrenia: a cognitive model of immunity to error through misidentification. In: Zahavi D (ed) Exploring the self: philosophical and psychopathological perspectives on self-experience. John Benjamins, Amsterdam/Philadelphia, pp 203–239

Gallagher S (1997) Mutual enlightenment: recent phenomenology in cognitive science. J Conscious Stud 4(3):195–214

Gallagher S, Cole J (1995) Body schema and body image in a deafferented subject. J Mind Behav 16:369–390

Gallagher S, Overgaard M (2005) Introspections without introspeculations. In: Aydede M (ed) Pain: new essays on the nature of pain and the methodology of its study. MIT Press, Cambridge, MA, pp 277–289

Gallagher S, Sørensen JB (2006) Experimenting with phenomenology. Conscious Cogn 15(1):119–134

Gallagher S, Zahavi D (2008) The phenomenological mind. Routledge, London

Graham G, Stephens GL (1994) Mind and mine. In: Graham G, Stephens GL (eds) Philosophical psychopathology. MIT Press, Cambridge, MA, pp 91–109

Husserl E (1970) Cartesian meditations (trans: Cairns D). Martinus Nijhoff, The Hague

Jack AI, Roepstroff A (2002) Introspection and cognitive brain mapping: from stimulus–response to script–report. Trends Cogn Sci 6(8):333–339

Jack AI, Roepstorff A (2003) Trusting the subject, vol 1. Imprint Academic, Exeter

Lutz A (2002) Toward a neurophenomenology as an account of generative passages: a first empirical case study. Phenomenol Cogn Sci 1:133–167

Lutz A, Lachaux J-P, Martinerie J, Varela FJ (2002) Guiding the study of brain dynamics using first-person data: synchrony patterns correlate with on-going conscious states during a simple visual task. Proc Natl Acad Sci USA 99:1586–1591

Le Van Quyen M, Adam C, Lachaux JP, Martinerie J, Baulac M, Renault B et al (1997a) Temporal patterns in human epileptic activity are modulated by perceptual discriminations. NeuroReport 8:1703–1710

Le Van Quyen M, Martinerie J, Adam C, Varela FJ (1997b) Unstable periodic orbits in a human epileptic activity. Phys Rev E 56:3401–3411

Le Van Quyen M, Martinerie J, Baulac M, Varela FJ (1999) Anticipating epileptic seizure in real time by a nonlinear analysis of similarity between EEG recordings. NeuroReport 10:2149–2155

Le Van Quyen M, Martinerie J, Navarro V, Boon P, DeHave M, Adam C et al (2001a) Anticipation of epileptic seizures from standard EEG recordings. Lancet 357:183–188

Le Van Quyen M, Martinerie J, Navarro V, Baulac M, Varela FJ (2001b) Characterizing the neuro-dynamical changes prior to seizures. J Clin Neurophysiol 18:191–208

Le Van Quyen M, Petitmengin C (2002) Neuronal dynamics and conscious experience: an example of reciprocal causation before epileptic seizures. Phenomenol Cogn Sci 1:169–180

Marcel AJ (1998) Blindsight and shape perception: deficit of visual consciousness or of visual function? Brain 121:1565–1588

Merleau-Ponty M (1962) Phenomenology of perception (trans: Smith C). Routledge and Kegan Paul, London

Nielsen TI (1963) Volition: a new experimental approach. Scand J Psychol 4:225–230

Noë A [ed.] (2007) Special issue on heterophenomenology. Phenomenol Cogn Sci 6(1–2)

Overgaard M, Ramsøy T, Gallagher S (2008) The subjective turn: towards an integration of first-person methodologies in cognitive science. J Conscious Stud 15(5):100–120

Paillard J (1999) Body schema and body image: a double dissociation in deafferented patients. In: Gantchev GN, Mori S, Massion J (eds) Motor control, today and tomorrow. Bulgarian Academy of Sciences, Academic Publishing House, Sofia, pp 197–214

Paillard J (1997) Divided body schema and body image in peripherally and centrally deafferented patients. In: Gurfinkel VS, Yu S. Levik (eds) Brain and movement. Institute for Information Transmission Problems RAS, Moscow, p. 141

Petitmengin-Peugeot C (1999) The intuitive experience. In: Varela F, Shear J (eds) The view from within. Imprint Academic, Exeter, pp 43–77

Price DD, Aydede M (2005) The experimental use of introspection in the scientific study of pain and its integration with third-person methodologies: the experiential-phenomenological approach. In: Aydede M (ed) Pain: new essays on the nature of pain and the methodology of its study. MIT Press, Cambridge, MA, pp. 243–274

Roepstorff A, Jack A (2004) Trusting the subject, vol 2. Imprint Academic, Exeter

Roy JM, Petitot J, Pachoud B, Varela F (1999) Beyond the gap. An introduction to naturalizing phenomenology. In: Petitot J, Varela F, Pachoud B, Roy JM (eds) Naturalizing phenomenology. Stanford University Press, Stanford, CA, pp 1–80

Schooler JW (2002) Re-representing consciousness: dissociations between experience and metaconsciousness. Trends Cogn Sci 6(8):339–344

Sørensen JB (2005) The alien-hand experiment. Phenomenol Cogn Sci 4(1):73–90

Stephens GL, Graham G (2000) When self-consciousness breaks: alien voices and inserted thoughts. MIT Press, Cambridge, MA

Tsakiris M, Haggard P (2005) Experimenting with the acting self. Cogn Neuropsychol 22(3/4): 387–407

Varela F (1996) Neurophenomenology: a methodological remedy to the hard problem. J Conscious Stud 3:330–350

Varela F, Shear J (eds) (1999) The view from within. Imprint Academic, Exeter

Vermersch P (1994) L'Entretien d'Explicitation. ESF, Paris

A Toolbox of Phenomenological Methods

Daniel Schmicking

'Phenomenology': One Term – Many Meanings

In addition to a large number of misrepresentations of phenomenology,[1] one can see an increasing interest in phenomenology among cognitive scientists and analytic philosophers. It is the method of phenomenology from which one expects to shed some light on the problem of consciousness. How does this approach work? What are the specific tools of phenomenology? Explanations as well as critical discussions of phenomenological methods are scattered across the literature. Moreover, the sometimes misleading or impenetrable terminology of classical phenomenology blocks easy understanding. The language of phenomenology may be an important reason why the working cognitive scientist who may be ready to delve into the primary sources of phenomenology might be discouraged, puzzled or disappointed by the notorious difficulty of many of the classical phenomenological authors.[2] Thus the main goal of this essay is to offer a sketch of the methods of phenomenology, which appeals to outsiders of phenomenology, and, hopefully, to a few insiders

D. Schmicking
Johannes Gutenberg University, Mainz

[1] I am using the term 'phenomenology' in a technical sense, referring to the philosophical movement that originated with Edmund Husserl, throughout the chapter. A clarification of the term will follow immediately.

[2] Phenomenological terminology is notoriously tricky. Many phenomenologists regard the language of classical phenomenology as often misleading or as Robert Sokolowski has put it 'a handicap for the phenomenological movement' (Sokolowski 2000, 226). Given their objective of interdisciplinary discourse or even collaboration with cognitive scientists, some phenomenologists are not content to retain problematic terms and thus give them up. Abandoning certain terms does, of course, not mean that phenomenological insights are also given up. If a phenomenological term is misleading and/or can be explicated by means of a more recently established, interdisciplinarily amenable term, the traditional term should be substituted. Thus, instead of trying to bend, say, Husserl's terminology to current purposes we may substitute terms like 'association', 'passive synthesis', 'eidos', 'monad' or juxtapose the more useful of them and novel terms that explicate the traditional ones. Hence one of the future tasks of phenomenology will be the clarification and major overhaul of its terminology that currently makes it difficult for scientists as well as analytic

too, if it can provide a (not quite) new way of looking at some (not quite) old ideas. After a short remark on the general character of phenomenology (Part II), phenomenological methods will be presented as a series of steps (Part III) and as a toolbox (Part IV). A concluding remark relates the offered account of method to the issue of the naturalization of phenomenology (Part V).

It is important to make clear, at the outset, how 'phenomenology' is used throughout this chapter (and most papers in this volume) because there are a bewildering number of meanings associated with this term today. The term serves to refer not only to diverse philosophical and scientific approaches but also to the subject matter of various accounts. A rough idea will suffice for our purposes here.

(1) We find the term used as broadly synonymous to what many authors in cognitive science and philosophy of mind call 'subjective experience,' 'qualia,' the 'what it is like' or 'phenomenal consciousness', i.e., the thing under study. (2) Further, there are various approaches in psychology, psychopathology, and psychiatry that have been called 'phenomenological'. (3) We should mention, at least, the many uses of the term in the history of philosophy. The first philosopher to effectively use the term 'phenomenology' (*Phänomenologie*) was J. H. Lambert in 1764, followed by Kant, Hegel, Lotze, Hartmann, Hamilton, Peirce, and, of course, Brentano among others.[3] (4) More central to the concerns of this volume, there is the philosophical movement that has developed from the early twentieth century on, in the writings of Husserl, Heidegger, Merleau-Ponty, and so on. (5) More recently there have been a number of uses made of the term in the field of cognitive science. Emerging from his critique of classic phenomenology (or what he deems it to be) Dennett, for example, dubs the method he favors 'heterophenomenology'. Following use (4), Varela (1996) proposes the term 'neurophenomenology' to signify the use of phenomenological methods in combination with neuroscience.

It is not always immediately clear in which sense authors use 'phenomenology'. In these cases, only the context disambiguates the use of the term. What makes things confusing is the way terms sometimes are 'hijacked' by theorists, and their meanings thereby loosened or made ambiguous (cf. Gallagher and Zahavi 2008, 13). In this chapter, and for the most part, in this volume, the term will be used as in (4). Here is a definition to give you a first idea of 'phenomenology'.

---

philosophers to understand correctly the classical phenomenological texts or recent work written in phenomenological argot. I have tried to find a middle way, i.e., I substitute some of the more cumbersome or misleading terms by less 'metaphysically laden' terms or I add the traditional term after referring to phenomena by recent terms. – I am fully aware of the perils of this practice. For instance, recent terms taken from psychology and cognitive science presuppose or imply a naturalistic stance. I put up with possible misinterpretations though, since I attach particular importance to the prospect of expanding interdisciplinary communication.

[3] The first verifiable use of the adjective '*phänomenologisch*' is found at least as early as 1762 in Oetinger Cf. Spiegelberg (1994, 11–19), Moran (2000, 6–9).

> 'Phenomenology' is, in the twentieth century, mainly the name for a philosophical move-
> ment whose primary objective is the direct investigation and description of phenomena as
> consciously experienced, without theories about their causal explanation and as free as
> possible from unexamined preconceptions and presuppositions (Spiegelberg 1975, 3).

Nota bene, Spiegelberg speaks of a movement, not a school, or a method.
Phenomenology is not monolithic, as we will see in due course. Its methods, how-
ever, can aptly be explained as a set of steps or tools. Particular combinations of
tools often correspond to various manifestations of phenomenology.

There is one further point that deserves particular mention. My remarks on
phenomenological method are primarily based on Husserlian phenomenology
and Spiegelberg's account of the phenomenological method. This may seem
biased to some readers. However, I do not regard Husserl's phenomenology as
the phenomenology. The tools I am going to present exceed Husserl's
approach, at least, if understood in the strict sense. Moreover, there are good
reasons why a presentation of the methodology of phenomenology is orien-
tated to Husserl. The most obvious reason is the historical fact that all phe-
nomenologists (no matter whether they have followed or criticized Husserl)
arrived at their own versions of phenomenology by taking their starting point
in Husserl. But beyond the historical dependence there is a methodic reason.
Here is how a phenomenologist, who has gone far beyond the Husserlian
framework, has put it:

> […] while Husserl cannot have the last word about phenomenology, he must have the first
> word. The analytic mind at work in the Husserlian style of phenomenology must inform all
> phenomenological study and distinguish between phenomenology proper, hard-headed
> phenomenology if you like, from soft-headed imitators. Husserlian phenomenology moves
> step by step, makes fine distinctions, and solidifies each item before moving on to the next
> development (Ihde 1977, 119).

Phenomenology – Just 'a Way of Seeing'?

Despite the efforts and achievements of Husserl and several of his co-workers
and followers to establish phenomenology as a program of philosophical
research phenomenology did not develop into a single coherent system or doc-
trine. Therefore phenomenology has aptly been characterized as a movement or
current of philosophical thought. The philosophers who consider themselves
part of this movement do not share a generally accepted set of methods and theo-
retical assumptions. There is, however, a widely accepted way to characterize
phenomenology: it is said to be 'a way of seeing rather than a set of doctrines'
(Moran, *Introduction* in Moran and Mooney 2002, 1). Description lies at the
heart of this way of seeing. Thus phenomenological description is regarded
as inseparable from phenomenology or as something like the least common
denominator of all varieties of phenomenology. For instance, Heidegger even
qualifies the expression 'descriptive phenomenology' as 'at bottom tautological'

(Heidegger 1962, 59). Many phenomenologists have made statements to this effect. Thus (1) it is widely agreed to identify phenomenology with a form of description to a greater or lesser extent. (2) There is a strong tendency to regard the objects of phenomenological descriptions as already known (phenomenology uncovers the sense of the world, which the world has prior to analysis). (3) Thereby the importance of theory for phenomenology is underestimated, ignored or even denied. True, phenomenology is *primarily* a descriptive discipline that does not offer, for instance, causal scientific explanations. More often than not, phenomenological analyses start with a 'describe, don't explain'. Theory does enter later in a good deal of phenomenology. Yet, repeated assertions that phenomenology is nothing but description or a way of 'seeing' are apt to entice people both outside and inside of phenomenology to regard it as hostile to theoretical explanation, and as confined to the realm of first-person experience and some kind of *elucidation*: 'the bringing to concepts of something we (in some way) already know, rather than the attainment of or claim to new knowledge of some phenomenon' (Glendinning 2007, 16).

Little wonder then, theoretically-minded authors such as Dennett insist on their story. Phenomenology, in particular, Husserlian phenomenology, according to Dennett, is 'based on a special technique of introspection …, the activity of introspection is a matter of just looking and seeing.…What we are fooling ourselves about is the idea that the activity of "introspection" is *ever* a matter of just "looking and seeing".… we are always actually engaging in a sort of impromptu *theorizing* …' (Dennett 1991, 44, 67).

It should become clear, in the course of this chapter, that phenomenology is neither just a matter of looking and seeing (thus limited to some kind of elucidation, for that matter) nor doomed to impromptu theorizing. Even if a great deal of what phenomenologists describe is – implicitly, "somehow" – known to us prior to phenomenological analysis it is possible to use phenomenology as a tool to discover new phenomena, experiences and structures that are not known prior to analysis. Moreover, there are theoretical foundations that phenomenology shares with other disciplines or philosophical approaches. Dennett's last point is very interesting in this respect. He claims that (a) phenomenologists fool themselves about the nature of their allegedly presuppositionless seeing, and (b) that there is always some spontaneous theorizing involved, of which phenomenologists supposedly are not aware. Contrary to official assertions, in Husserl's phenomenology, which is Dennett's main target, there is, for instance, a theory of formal ontology presupposed by Husserl's descriptions and analyses, which certainly is no case of spontaneous or provisional theorizing. Another example is the layered theory of mind that one finds in Husserl's genetic phenomenology, Merleau-Ponty and in current theories developed by phenomenologists. The dependencies and implications that are the subject matter of genetic phenomenology are forms of phenomenological explanation (cf. Sass, this volume). One finds similar layered theories in developmental psychology, psychiatry and neuroscience (e.g., A. Damasio, D. Stern, C. Trevarthen).

Spiegelberg's Account of Phenomenological Method as a Series of Steps

The huge number of publications on and in phenomenology notwithstanding, it is not easy to find texts that deal with the *method* of phenomenology head-on and in an accessible way.[4] A very clear and helpful text in this respect is Spiegelberg's "The Essentials of Method" (Spiegelberg 1994, Part 5), from which I draw the starting point of my considerations. The decisive advantage of Spiegelberg's approach is the fact that it is written from the point of view of the active phenomenological researcher rather than from an historical-exegetical perspective.

Spiegelberg characterizes the phenomenological method as a series of steps. Elsewhere, Spiegelberg used the simile of a core around which the varieties "can be arranged like concentric shells" (Spiegelberg 1975, 56) to illustrate the relationship of the versions of phenomenology to their common methodological foundation.

The concept of a series of steps not only embraces core methods used in phenomenology but can also explain many of the differences and relations between various types of phenomenology. The steps of the phenomenological method according to Spiegelberg (1994), 681–717) are:

1. Investigating particular phenomena
   - Intuiting
   - Analyzing
   - Describing
2. Investigating general essences
   - Intuitive apprehension
   - Analysis
   - Description
3. Apprehending essential relationships among essences
   - Internal relations within one essence
   - Relations between several essences
4. Watching modes of appearing
   - Sides/aspects
   - Perspective shading-offs (*Abschattungen*)
   - Modes of clarity
5. Watching the constitution of phenomena in consciousness
6. Suspending belief in the existence of the phenomena
7. Interpreting the meaning of phenomena

(1) The first step of this list is what most authors refer to as 'description'. Actually it involves a series of three component steps, or phases. The very first activity is called 'intuiting' by Spiegelberg. Probably, it is this activity that

---

[4]One has to mention, at least, Ihde 1977 and Embree 2007 here. The latter probably is the only real hands-on introduction to doing phenomenology, which is apt for both self-study and use in groups (the classroom).

particularly motivated the widely accepted characterization of phenomenology as a 'way of seeing', unlike theories or doctrines.

In the 'intuitive' grasping of phenomena[5] under consideration one tries to experience things in a way that is "as free as possible from unexamined preconceptions and presuppositions" (Spiegelberg 1975, 3). This kind of opening and grasping means among other things that all details have equal rights. Nothing is ruled out or considered unimportant because some established view has it that way. As opposed to the mere grasping of phenomena, the next phase includes *analysis*, which consists mainly in distinguishing the constituents of the phenomenon and the relations among them and to adjacent phenomena (cf. Spiegelberg 1994, 691). *Describing* the structure that emerges from the analysis is the final task of the first step. Classical phenomenologists such as Husserl and Spiegelberg have emphasized that the preceding phases of grasping and intuiting are a matter of pre-predicative experience. 'Phenomenology begins in silence. Only he who has experienced genuine perplexity and frustration in the face of the phenomena when trying to find the proper description for them knows what phenomenological seeing really means' (Spiegelberg 1994, 693). Let us ignore the conceptual and linguistic problems involved here because they do not significantly differ from other fields or disciplines, e.g., words often are not fine-grained enough to report perceptual nuances exhaustively; words carry connotations that may distract readers or are inapt to denote new structural features that phenomenological analysis has discovered etc. If there are no adequate general terms to classify particular phenomena because they are new or looked upon from a new point of view one may resort to metaphors and analogies etc.

An important function of phenomenological description (at this stage) is to serve as a signpost to the phenomena. Understanding a phenomenological description means for the readers, first and foremost, to experience themselves the reported structures. (For an inchoate analysis you need not finally decide on the terms you give the components of the structure you have arrived at. You try to make the readers perceive or imagine what you had grasped).

(2) The second and the third step are likely to raise grave suspicion among scientists, philosophers in general, and even phenomenologists. While many descriptions start from particular phenomena (for instance, Husserl's notorious inkpot), phenomenologists aim at insights in the general structures of objects and experiencing. The step they take from particular cases to universal statements about essential or apriori structures requires a particular tool, which, in the second step, makes it possible to determine 'essences,' or, to use a less suspect word, 'invariant structures'. Spiegelberg opines that this phase again involves an intuitive grasping, analyzing and describing. He distinguishes this step from the third step, which comprises the determination of the internal relations within one essence, and the relations between different essences. It is important to note that, according to Spiegelberg,

---

[5] The term 'phenomenon' indicates that the respective object is considered from the phenomenological perspective. More specifically, depending on the methodological framework, the term 'phenomenon' may indicate that the object under consideration is phenomenologically reduced (see step 6, below).

only this step requires free imaginative variation. (I shall get back to this point in due course.)

Phenomenological analyses do not determine or discover essential or invariant structures by conceptual analysis. Phenomenology studies 'the things themselves' and the way they appear in experience. This way of appearing depends on a stronger (e.g., formal-ontological) or weaker (e.g., phylogenetic) apriori, namely the 'invariant structures' of the types or species the appearing things are members of (for instance, auditory or visual percepts, spatiotemporal objects, mathematical objects etc.). This holds for perceptual experience as well as experience involved in, say, deductive reasoning or speech acts.

(3) In Spiegelberg's definition of phenomenology quoted above, the primary objective of phenomenology is characterized as the 'direct investigation and description of phenomena *as* consciously *experienced*' (Spiegelberg 1975, 3, my emphasis). This may sound easy or natural but it poses a real challenge because the particular ways in which we experience things are hardly ever what we are directed at in everyday life. Quite the contrary, the typical modes of appearance are part of the basis of our habitual world. They are taken for granted, part and parcel of the ways we relate to situations, not what we pay attention to. (Note that 'appearance', in phenomenology, does not denote something like a merely appearing veil of reality as opposed to an objective reality-in-itself but just the various ways we experience things themselves. This holds for perceptual objects, which appear in spatial-perspectival, temporal and attentional modes, as well as for any kind of object such as remembered objects, deductive arguments or musical compositions.)

These modes are what phenomenology investigates in the fourth step. Structures, rules, and constraints of how we experience things in perception, imagination, conceptual understanding etc. are thus revealed. Phenomenology abstracts rules and constraints from the descriptions of the modes of appearance.

In studying the types of such modes and their dependencies, phenomenology is, among other things, concerned with what are called 'perceptual constancies.' This term refers to the way a constant perceptual property of an object is perceived as unchanging through a variety of proximal stimuli. For instance, a whistle appears louder when the vessel approaches but, at the same time, you know that this is due to the decreasing distance of the sound source. Husserl repeatedly deals with various perceptual constancies that, during his time, had not been studied experimentally. Taking up Husserl's observations, I have offered a systematic study of auditory constancies, some of which have not yet been mentioned in the literature (Schmicking 2003, §§26–29). This is only one example that may prove the practicality of a phenomenological approach. Moreover, it contradicts those views of phenomenology that doubt or deny the potential of phenomenological analysis to discover new phenomena. Thus phenomenology can be used as a valuable heuristic tool to determine new experimental questions and to design experiments not only in the neurosciences but in psychophysics (see Horst, this volume).

(4) Constitution is the subject matter of the fifth phase according to Spiegelberg's explanation. This term refers to the invariant structures of how modes of appearance

are interrelated (e.g., intermodally) and how, as it were, from these building blocks, revealed by the preceding step, things are compounded, i.e., how their (nonlinguistic, nonconceptual) meaning as objects of a certain type emerges from our interacting with the world and with others. Spiegelberg's heading ('constitution of phenomena in consciousness') and his short explication sound unnecessarily consciousness-centered (though his example is not, see Spiegelberg 1994, 707f.) Constitution is not an exclusively mental achievement: our bodies are inseparably involved in constituting the world. Phenomenology studies the ways various cognitive (mental and bodily) operations (including, of course, emotional processes) combine to constitute an object as having a meaning for human beings. Objects – whatever kinds of objects you assume there are: spatiotemporal, mental, social, ideal etc. – 'are not simply *in* consciousness as in a box' (Husserl 2001, I, 275). Nothing evades constitution, not even consciousness itself (i.e., there are invariant structures of consciousness). Husserl explicitly states that nature (the lived) body and "pure consciousness" constitute themselves in mutual dependency (Husserl 1952, 124).

(5) The sixth step, suspending belief in the existence of objects, is often involved in the first step, i.e., grasping or discovering phenomena 'as free as possible from preconceptions and presuppositions'. If you suspend your (scientific or common-sensical) belief in the existence of objects and events, you are more likely to reach a minimum of preconceptions and presuppositions. This suspending of belief is the notorious *phenomenological reduction*. The reduction is neither indispensable nor undisputed however. Phenomenology without the reduction is possible, and although Husserl is the 'inventor,' and most important advocate of the reduction, some of his earlier work is carried out without using this tool.[6]

The basic idea is as follows: The more you are engaged in a situation, the less you are aware of all constituents of the situation, the less objective you are in your account of the situation, or rather, you cannot give an account while 'absorbed' in the situation at all. The function of the reduction is to step back from the situation (i.e., ultimately from the world, which is the all-encompassing horizon of all possible situations). The 'bracketing' of the existence of the world is simply an aid to make possible the maximally unprejudiced investigation. It is important to see that the suspending does not mean that the existence of the objects under consideration (or ultimately the world) is *denied*. The reduction serves to draw our attention to objects, events and mental structures such as Husserl's *Urdoxa* or Searle's default positions. The reduction allows us to conceive objects and events with the meanings they have for us, while we stop taking 'mundane' interest in the objects under consideration. The reduction is not an epistemological tool however, at least not in a sense that originates from the Cartesian–Lockean tradition, because its objective is not to verify or falsify (beliefs of) the natural attitude. As phenomenologists carrying out the reduction,

---

[6]This is the reason for Spiegelberg's 'manifest 'heresy'' of dealing with the reduction almost at the end of the series of steps (cf. Spiegelberg 1994, 708).

> [w]e must leave everything as it was, for otherwise we would change the very thing we
> wish to examine. […]. It [the reduction] does not try to substitute its own perspective […]
> for the perspectives of the natural attitude. It does not become imperialistic and does not
> claim that its mode of truth is the only one there is (Sokolowski 2000, 191).

There is a certain danger, though, that phenomenology never finds its way back to
questions of being. "Reductive phenomenology is no substitute for epistemology. Nor
can it ultimately become a substitute for metaphysics. There is no escape from the ear-
nestness and persistence of the question of reality and of being" (Spiegelberg 1994, 711).
The persistence of metaphysics is palpable in recent debates on consciousness, the body-
mind problem, free will, etc. If phenomenology wants to have a hand in those debates it
cannot abstain from ontological commitments once and for all. Maybe this is one reason,
among others, for the appeal of existential phenomenologists such as Heidegger or
Merleau-Ponty. For instance, the latter moderately criticized Husserl's tendency to
demand an absolute reduction. Merleau-Ponty interprets Husserl's essences as 'destined
to bring back all the living relationships of experience …' (Merleau-Ponty 2002, xvii).
The reduction discloses our situation as being-in-the-world, according to this understand-
ing, which is much closer to the late Husserl than the bulk of the phenomenological lit-
erature suggest (see, e.g., Thompson (2007, p. 21f.) for a judicious appraisal).

(6) The seventh step is concerned with more implicit or veiled aspects of being-in-
the-world, which are mainly studied by *hermeneutic* or *existential* phenomenology.
Here is a very rough idea of it: The goal of existential phenomenology is to disclose
and interpret the meaning or sense of phenomena that are hidden, 'not immediately
manifest to our intuiting, analyzing and describing' (Spiegelberg 1994, 712), i.e., phe-
nomena that allegedly lie beyond the scope of descriptive phenomenology. Those
hidden meanings are, for instance, care (*Sorge*) in Heidegger's hermeneutic of human
being, indicated, for example, by being in a certain mood, or Sartre's 'deciphering' of
the human being's 'existential projects'.

Spiegelberg voices two concerns about the hermeneutic approach: First, many
have asked whether these kinds of analysis add a new tool to phenomenology.
Further, while it is difficult enough to find ways to counter-check descriptive
phenomenology, it is far from clear how the statements of hermeneutic phenome-
nologists can be made objective, or intersubjective, or valid.

The issue whether existential (or hermeneutic) phenomenological analyses differ
substantially from 'merely' descriptive phenomenology is a difficult one, which would
deserve a detailed discussion. Here I can only hint at the direction that my explanation
would take (and which I just indicated in the preceding remark on the reduction).

There is a broad consensus on contrasting transcendental (Husserlian) with exis-
tential phenomenological approaches. Existential or hermeneutic phenomenologists
claim that transcendental approaches fall prey to what Heidegger calls "a basic
deception for phenomenology,' namely 'having the theme [of description] deter-
mined by the way it is phenomenologically investigated' (Heidegger 1985, quoted
from Wrathall 2006, 41). By allegedly mistaking the theoretical stance as the
'exemplary mode of being-in-the-world' a transcendental (Husserlian) phenome-
nology (of consciousness) is said to 'miss the pre-reflective, pre-conscious mode of
being in the world' (Wrathall 2006, ibid.).

Even if there is a tendency in some authors, notably Husserl, to focus on "a specifically theoretical comportment" (Heidegger, ibid.) one must not identify the theoretical stance of the reduction with the worldly experiencing or comportment that those authors investigate. Every phenomenologist – existential as well as transcendental – must engage in acting (perceptual, emotional, scientific, communicative comportment) to get started on intuiting and analyzing (i.e., the first step of phenomenological investigations). Phenomenologists have to zigzag between naïve worldly comportment and the theoretical stance. (Cf. the first remark on description in (3) below.)

True, the later Husserl tends to make his phenomenology seem like a basically idealistic position. However, if one looks out for what Husserl does rather than for his official self-portrayals one easily finds an abundance of studies in prereflective phenomena and prereflective intentionality ('primary and secondary passivity', in Husserl's terms). As existential phenomenology centers on these there is sufficient common ground for both to meet on. Transcendental as well as existential phenomenology start from a relationship (*Korrelationsapriori*, being-in-the-world), which is a continually developing, dynamic structure of interdependencies of self, others, and world. Husserlian transcendental phenomenology, when we do not misread it according to idealistic interpretations, already offers the tools necessary to study this relationship.

Phenomenological Methods as a Toolbox – Complementing Spiegelberg's Steps

In what follows, I will offer a slightly modified view of the methods of phenomenology, which complements the steps and explanations Spiegelberg puts forward and which considers more recent developments in phenomenology that interface with the cognitive sciences.

(1) Generally, I propose to speak of 'tools' rather than of 'steps'. The latter term seems to suggest a fixed order of the components, though one can skip one or more steps. However, often several of the tools are used simultaneously; different combinations and orders are possible and customary too. Consequently, phenomenological approaches differ in their choices of tools from the full range.

Just like hand tools the phenomenological tools are comparably independent of each other. If you fix or construct something in your apartment you probably do not use all of the tools from your toolbox, let alone all of them simultaneously, although combinations of two or more tools are common, e.g., a water-level and a pencil or screw clamps, a wood drill and a hammer to drive in a wooden dowel. Moreover, there are typical orders of using certain tools. (You happily use the water-level before you drive the nails into the wall.) Phenomenological investigations work in a similar fashion. You engage in a description of phenomena typically by intuiting and analyzing them while suspending commonsensical and scientific explanations, i.e., taking two steps or using two tools at the same time: reduction and description. You continue, for instance, using formal-ontological categories and (imaginative

and factual) variation to distinguish and classify the constituents of the type of phenomenon under consideration in an 'eidetic analysis'. The latter tool necessarily presupposes the preceding description of particulars.

(2) Spiegelberg's step 6 (the phenomenological reduction or *epoché*, which originally refers to the attitude that results from the reduction) does not contain a distinction of possible types or rather directions of suspending, though it makes quite good sense to distinguish between them. Spiegelberg's step 6 can thus be subdivided into three stages or directions of reduction, namely (a) suspending of commonsensical and scientific beliefs (assumptions of naïve physics and psychology, scientific causal explanations etc.), (b) suspending of belief in the existence of objects/world, and (c) suspending of the belief in an extended biographical self. The latter *transcendental* reduction in the narrow sense brackets everything that is not part of the present mental content, for instance the past moments of one's own life.

The phenomenological reduction is the distinctive move from the natural attitude, i.e., from the 'default mode of our acceptance of the world and things in it' (Sokolowski 2000, 45), to reach the reflective stance of phenomenology. By suspending one's beliefs in causal explanations, possibly also in the existence of the objects, and even in one's biographical self one is most likely to reach descriptions that are 'as free as possible from unexamined preconceptions and presuppositions' (Spiegelberg 1975, 3). This is the reason why the suspending of belief should rank first in the phenomenological toolbox, although many phenomenologists have been critical of the possibility of the phenomenological reduction. Distinguishing certain scopes of the reduction as indicated above might help to forestall misunderstandings, particularly concerning the alleged absoluteness of the reduction, and to render this tool more feasible.

(3) Concerning Spiegelberg's step 1, *description*, there are two additional points that deserve particular attention: What Husserl calls *Nachgewahren* is part and parcel of description, or rather of the descriptive phase which Spiegelberg aptly calls 'analysis'. The term *Nachgewahren* refers to the grasping of a lived experience immediately after the experience itself, i.e., while the content of this episode is still given in retention. Phenomenologists need to habitualize their interest in their subject matters (types of objects and correlative experiences). During performing everyday tasks, skills, etc. in the natural attitude the latent interest awakes immediately after certain experiences, acts, etc. In this way experience or worldly comportment can be analyzed without distortions, thereby avoiding the paradox of self-observation, which has been one of the main objections to self-observation in debates since Comte. Contrary to superficial notions of phenomenology, the practice of phenomenological 'seeing' requires a lot of time and patience. 'The apparent facility for access to one's experience is what is misleading: the filling-in [of the reflective experience directed on a previous content] made possible by suspension has its own developmental time which needs cultivation and to be borne patiently' (Varela and Shear 1999, 8)[.7]

---

[7]Cf. also Varela 1996 on the requirement of sustained training in phenomenology: 'This last aspect … is perhaps the greatest obstacle for the constitution of a research programme since it implies a disciplined commitment from a community of researchers' (op. cit., 338).

The second remark concerns the language of phenomenology. At early stages of description the function of how phenomenologists put things into words is often to guide the reader's imaginative experience so she can experience the 'things them-selves' under consideration. The language of phenomenology should be as free as possible from connotations, preconceptions etc. This is hardly achievable or realizable of course. Perhaps, the most successful attempt at a presuppositionless terminology is Heidegger's idiosyncratic terminology in *Being and Time* (Heidegger 1962). Moreover, phenomenological language should be as detailed (conceptually fine-grained) and as precise as possible. Symbolization and formalization are means of developing precise (exact, disambiguated) representations. However, only very few phenomenologists have tried to develop symbol systems or formalized languages as an exact medium of phenomenology. For instance, Eduard Marbach (1993, and this volume) has provided a phenomenological notation that allows representing various types of experiences, and relations of implication and dependency between different acts and their parts by means of a symbolism, which is inspired by Frege's *Begriffsschrift*. Other phenomenologists such as K. Mulligan, G. Null, B. Smith, and O. Wiegand have developed formalized or diagrammatical systems to represent mereological structures and to axiomatize phenomenological mereology. If these symbolic and formalized means were developed further and used by the majority of phenomenologists, a notation like Marbach's could essentially contribute to the validation of phenomenological analyses. A generally agreed upon notation might serve as an indispensable tool of cross-checking and possible agreement procedures in phenomenology, similar to the phonetic notation in descriptive linguistics.[8]

(4) Spiegelberg's Step 2 (intuiting general essences) is problematic, at least, as it stands. General essences or, carrying less metaphysical baggage, *invariant structures*, necessarily are structures that comprise abstract constituents (i.e., dependent and/or independent parts). It is not clear how one can 'intuit an essence' of, say, redness, color or physical force (Spiegelberg's examples) without analyzing the structure of redness, color or force (e.g., perceptual attributes) and relationships between the constituents of the structure. There is no mystical, single-step intuiting of essences. Instead of distinguishing two steps, investigating essences and investigating their relationships (Spiegelberg's step 2 and 3), I propose to lump these steps together into one tool: the *investigation of invariant structures and their relationships*. This tool is in turn dependent on another tool, formal ontology, which is a pure theoretic component of phenomenology. The importance of part/whole relations and of formal categories and dependencies in general has been widely underrated or neglected in phenomenology (for instance, Husserl's texts teem with 'foundation' (*Fundierung*) and '(real) moment' (*das (reelle) Moment*), the latter sometimes mistakenly rendered as 'instant'). However, formal ontological categories and relations form the conceptual framework that is presupposed by a great deal of the analyses of authors such as Husserl, Reinach, Pfänder, and is present in many others, even though they have not always applied this framework explicitly or as consequently as Husserl did.

---

[8]Cf. Schmicking (2005, 177–181) for a more detailed argument.

Moreover, contrary to some orthodox views, the analysis of invariant structures should be informed by empirical data (if available), thereby going beyond the 'free imaginative variation', which Husserl and classical phenomenology endorsed. Our imagination may take us far but it is not immune to one-sidedness, sometimes it is unaware of possible variations, and hence it might not always be 'free' enough.

While he always emphasized the necessity of eidetic analyses, Husserl realized that the latter cannot retrieve and determine *exact* essences (like those of pure mathematics) in the field of consciousness; here we find only *typical* forms, which Husserl calls *morphological*. There are, of course, exact and formal essences in the sciences such as mathematics or pure formal grammar. But it is important to see that 'merely' morphological essences are not a methodical shortcoming of phenomenology but due to the nature of its subject matter (cf. Husserl 1982, §§7, 10, 73, 74).

(5) Spiegelberg's steps neither present nor mention the theory of whole and part. However, this doctrine is a ubiquitous tool that is applied to all fields of phenomenology. Today there are a number of studies in mereology which build points of contact with analytic philosophy (Carnap, Goodman) and the formal sciences (mathematics and formal linguistics). Phenomenological mereology can be formalized and presented as an axiomatic system, which shows that there is a theory, which comes first or early in phenomenology. Phenomenology is not limited to 'impromptu theorizing'. There are, of course, further component theories in phenomenology, for instance, related versions of gestalt psychology and scene analysis (e.g., in Husserl 1988, Merleau-Ponty 1963) and a layered theory of mind (e.g., in Husserl 1973, 1988, Merleau-Ponty 2002), which anticipates current layered models of the self as developed by A. Damasio or D. Stern (see Fuchs, this volume).

(6) Spiegelberg's step 5, constitution, does not include the distinction between static and genetic constitution. It doesn't (explicitly) consider the role of the body in constitution either. I propose to split the step into two different tools, because often they are used independently, and to pay due attention to the body in both contexts. Here is an illustration.

Consider the constitution of a song as an object of your experience. Let's assume, a friend tells you about a song that you have not heard before. You start imagining (quasi-auditorily) the song based on auditory memories (e.g., of the singer's voice) and propositional knowledge (what is the song like: blues, r&b etc.). Next you might buy a CD that contains this song or download it. You listen to it for the first time: now you hear the melody, the lyrics, and you are aware of some instrumental parts. During repeated listening, you grasp the complete lyrics, your attention is drawn to further musical elements, e.g., the four bars intro, a guitar break etc. Let us further assume you want to play or sing the song by yourself. You try to play the song 'by ear,' you start to analyze the song by figuring out the chords on your guitar. You have difficulties so, finally, you get yourself a score of the song, practice the chords, and you sing along.

There are various levels or dimensions of experience and their correlates involved here: after the initial linguistic information, which first draws your attention to the object, you imagine quasi-auditorily what the song might sound like. Then you have auditory experience of it, joined by linguistic acts of understanding the

lyrics. After listening to the song you voluntarily or involuntarily remember parts of the song (maybe it is a catchy tune). Further you add sensorimotor experience, and finally, using a visual symbolic representation of the melody and the chords, you gain additional abstract, categorically articulated knowledge of the musical structure of the song (this time more comprehensive than when you fumbled with your guitar). This example is, of course, merely one possible individual trajectory of constituting a song as experienced. Constitutional trajectories are different for every subject and every constituted object. Again, phenomenology necessarily must take the step from individual examples to essential structures. It discovers general rules and constraints of constitution, e.g., gestalt laws of scene analysis in visual and auditory perception, or 'foundational' (i.e., dependency) relationships between the different types of symbolic, perceptual and imaginative (re)presentation etc.

A *static* phenomenological analysis determines the various types of experience (parts of the invariant structure called 'transcendental subjectivity') and its objects; in the example the types that are involved in how we learn a song, musical categories etc. A *genetic* analysis studies how the cognitive structures and the respective types of content develop through time. For instance, reading (i.e., understanding) scores presupposes prior knowledge of a tonal system and a musical syntax, as does auditory imagination. Initial rule-following, categorically articulated movements, e.g., playing chords on the guitar, habitualize until they have 'sunk down' on a prereflective level of motor intentionality. (This way of automatization is called 'secondary passivity' by Husserl.)

Thus particular trajectories are only starting points of constitutional investigations. Individual trajectories of experiences must be varied eidetically, i.e., they are varied imaginatively as well as based on empirical variations (if available) until one reaches the invariant elements of constituting a type of object or experience.

Phenomenology of constitution studies the ways various cognitive (mental-affective-bodily) operations combine to constitute an object as having a meaning for human beings. Objects – whatever kinds of objects your ontology assumes: physical, spatiotemporal, mental, social, ideal etc. – 'are not simply *in* consciousness as in a box' (Husserl 2001, I, 275). There are 'a priori limited ways of' how they can 'appear' to us, and how the objective (*gegenständlich*, i.e., not linguistic) meanings of things develop, how those meanings change, according to new experiences, or sometimes 'explode'. Nothing eludes constitution, not even so-called pure consciousness itself.

There are two further tools of phenomenology which are of particular importance for current developments in phenomenology.

(7) Analyses of isolated phenomenologists who investigate the same objects and who even cross-check their results seem to be limited to first-person experience – even if their findings turn out to be consistent. Contrary to a widely accepted view, phenomenology is not necessarily confined to subjective or private experience (i.e., limited to the narrow window of the subject's living present and memories, if the latter are considered as reliable). There are various ways to widen the range of phenomenological data, though.

(a) One way of crossing the border is the use of data including statistical evidence as well as case studies, private reports etc. from psychology, psychophysics,

cognitive neurosciences, clinical neurology, psychopathology etc. Phenom-
enologists can thus make use of empirical data that is produced in the mundane, natural
attitude of the sciences either without carrying out the phenomenological reductions or
'reducing' the data back into their transcendental attitude (see Schmicking (2003),
who makes extensive use of psychoacoustics). Further, Husserl occasionally consid-
ers bodily or mental dysfunctions or disabilities, e.g., cases of war invalids, perturba-
tions of perception caused by drugs such as Xanthonin, etc.; Merleau-Ponty, however,
makes systematic use of mental and bodily dysfunctions. He considers case histories
from psychiatry, psychopathology, physiological, behavioral, and subjective data to
develop a phenomenological understanding of the way the embodied mind works. His
approach is thus closely related to the *lesion method* in cognitive neurosciences, which
is still deemed to be the gold standard, according to Hannah Damasio (Damasio 2001,
109). The phenomenological interpretation of experimental data and explanations
leads to what Gallagher (2003) has called *indirect phenomenology*. The interpretation
of dysfunctional data in Merleau-Ponty is a first systematic manifestation of this
approach. We may distinguish the use of experimental data and explanations from
(case) studies of dysfunctions and studies of normal functions, respectively. By using
those kinds of data phenomenologists can complement and correct, for instance, their
limited 'free' imagination on which they rely in investigations of invariant structures.

(b) There is another interesting option, which still seems to await systematic
development: phenomenologists could design and carry out their own experiments,
i.e., within the phenomenological attitude. Husserl himself proposed to conceive an
apt experimenting that should not be 'inductive-objective' (*induktiv-objektiv
gerichtetes Experimentieren*) (cf. Husserl 1988, 150f.).

Hence we should, at least, add a tool to Spiegelberg's list, which we might call
interpreting experimental (dysfunctional) data. What has recently been developed
and aptly called *frontloaded phenomenology* is a compound method: the initial
phenomenological data or testable hypothesis is obtained by applying phenomeno-
logical tools to certain phenomena (with an eye to particular techniques or issues in
cognitive neurosciences) and then handed over to the neuroscientist. The experi-
mental results, in turn, will not only be interpreted in the framework of neurosci-
ences but also from a phenomenological point of view, and, moreover, can
provide criteria to correct or modify the original phenomenological analysis.
Neurophenomenology is a closely related compound method, which integrates tools
of classical phenomenology and neurosciences but differs in that it requires sub-
jects to be trained in, at least, fragments of phenomenology (cf. Gallagher 2003 and
this volume; Gallagher and Sørensen 2006; Varela 1996).

(8) As the preceding tool shows, phenomenology is not limited to so-called first-
person private experience. For instance, there are some older attempts at integrating
vicarious experience into phenomenological studies (Jaspers, Binswanger), and
recently there are researchers that use various interview techniques which are based
in phenomenology as well as in empirical psychology (e.g., Petitmengin 2006), and
sometimes even go back to introspective psychology. For instance Vermersch
(1999) confidently seeks to improve introspection as a practice by passing from a
first-person to a second-person point of view. Among other things, these novel

second-person approaches have shown that the often taken for granted demarcation line between the conscious and the non-conscious is far from clear or fixed. The notion of a simple split (conscious vs. non-conscious/sub-personal/etc.) must be abandoned. Husserl's conception of secondary passivity already has suggested that there is no 'once-and-for-all' fixed boundary between what is conscious to the reflective mind and what is not. Depending on one's activities, learning etc. there is a whole bundle of lines which are likely to be domain-specific and in a continual change for every skill for every individual respectively. For instance, certain lines move according to the increasing automatization of skills. Novel methods such as Petitmengin's allow for the uncovering of fields of pre-reflective processes. Those investigations can contribute to an improved, widened view of the margins of the 'conscious' and 'non-conscious' and the vast area between the cogito and bare physiology, which is hardly explored in the neurosciences yet. Further, second-person methods allow for validation and evaluation more easily.

Finally, second-person methods illustrate a further point I have not mentioned as yet. Phenomenologists are toolmakers as well as users of, so to say, prefabricated tools.[9] Thus some of the tools that I have just mentioned still are in the making. Probably there are further possible tools for phenomenology that are not even in the design or planning stage. In general, phenomenology ought to strive to overcome solitary reflection. Recent interdisciplinary approaches such as neurophenomenology and front-loaded phenomenology are ways to overcome the relatively isolated situation of the working phenomenologist, which moreover allow for (more or less mediate) evaluation of phenomenological results.

Given these considerations we reach at the following list of phenomenological tools:

1. Phenomenological reductions
   - Suspending commonsensical and scientific explanations
   - Suspending belief in the existence of objects/world
   - Suspending belief in extended biographical self
2. Investigating particular phenomena ('description')
   - Detecting and grasping (*Nachgewahren*)
   - Analyzing
   - Describing (embracing common and technical language as well as symbol systems)
3. Mereological analysis
4. Investigating invariant structures and relationships
5. Analyzing typical/invariant modes of appearing
6. Analyzing the static (embodied/kinesthetic) constitution of objects/experience
7. Analyzing the genetic (embodied/kinesthetic) constitution of objects/experience
8. (Hermeneutic) interpretation of the meaning of existential phenomena

---

[9] Thanks to Brunce Janz for drawing my attention to this aspect.

9. Interpreting experimental (dysfunctional) data and explanations
10. Interpreting vicarious experience and second-person methods in general

Naturalization of Phenomenology – a Conciliatory Proposal

It should have become evident from the preceding sections that phenomenology is neither monolithic nor static. There is no single view of phenomenology and hence no single way of doing phenomenology. The various preferences or choices of tools show family resemblance to various degrees. Necessarily the pluralism of phenomenological approaches will affect the issue of naturalizing phenomenology.[10] In the following, I will argue that the decisive question is not whether phenomenology should be naturalized, as it were, 'neck and crop'. The crucial point is rather about the function of two particular manifestations of phenomenology and their relationships: transcendental phenomenology as first philosophy, and empirical, interdisciplinary phenomenology as naturalized philosophy.

Philosophy was once regarded as the ultimate medium of foundation or justification of all types of knowledge. Some of Husserl's official programmatic statements seem to aim at this ideal of first philosophy, for instance, in *Philosophy as a Rigorous Science* (Husserl 1964). Similarly, in his lectures on *First Philosophy* (Husserl 1959), phenomenology is conceived as a universal science, which claims absolute justification, and in the *Crisis*, Husserl regards transcendental phenomenology as the 'constitutive 'internal' method, through which all objective-scientific method acquires its meaning and possibility' (Husserl 1970, §55, 189). However, Husserl (as well as most other phenomenologists) never seeks for an absolute foundation in a traditional metaphysical sense. The presuppositionlessness that phenomenology strives for is embodied in the consistent application of the reductions. Consequently phenomenology does not take for granted the notions of being, causality or subjectivity (in any metaphysical sense of these terms). Further, Husserl has never endorsed doctrines of infallibility and incorrigibility, nor of a privileged access to the 'internal' sphere. Even if some of his statements suggest foundationalistic interpretations we should not ascribe any traditional form of foundationalism to Husserl or any sober phenomenologist. (For elaborate arguments to regard Husserl's position as non-foundationalist, cf. Drummond 1990, and Zahavi 2003). Yet there still is the legitimate phenomenological idea of first philosophy as a regulative idea.

I cannot address, in the limits of this essay, the manifold questions and problems that the notion of a 'naturalization of phenomenology' poses. First, there is no one phenomenology. Further, there are different, though related, senses of 'naturalism' and 'naturalization'. Thus both relata in the 'naturalization of phenomenology' bear a certain ambiguity. A clarification of this possible project raises many questions. Among others:

---

[10]As Dan Zahavi's essay (this volume) is dedicated to the issue of naturalization of phenomenology I limit my discussion to a few remarks that may complement Zahavi's argument.

- Is a naturalized phenomenology committed to the ontological assumptions of naturalism? If it is to which assumptions? Many phenomenologists will readily assume that human beings are part and parcel of the natural order. There are different options here. One is materialism (an option which most phenomenologists will dismiss as self-defeating). Another option would involve an opposite move some authors have called a 'phenomenologizing' of science. Roy et al. (1999) sketch out a strategy they call *naturalization as enlargement of the concept of nature*. This project goes back, at least, to the early Merleau-Ponty (1963). The enlargement consists in ascribing *form* (invariants, macroscopic properties) to physical nature. The decisive point is that the notion of form is not deducible from the physical world but only from the phenomenological domain (cf. Thompson 2007, ch. 4).
- Will a naturalized phenomenology assume that scientific data and theories have relevance for philosophical problems (and vice versa)? This is likely to be accepted by many phenomenologists today, and it would open strategies of mutual constraints for the program of neurophenomenology, as Varela (1996) has suggested.
- Will a naturalized phenomenology assume that its descriptions will be reduced to scientific propositions in that only propositions that are reducible to the terminology of (natural, cognitive, behavioural, social?) science are valid and justifiable? This would land phenomenology in a traditional reductionism or eliminativism, which probably all phenomenologists want to avoid.
- Finally, phenomenologists cannot be expected to assent to strong versions of scientism, which claim that only methods of the natural sciences are reliable or justifiable methods.

These few hints already show that there is no straightforward way to answer the question of how a naturalization of phenomenology could be carried out. Yet, we need a notion of a naturalized phenomenology, even if a provisional one. Here is my suggestion: If a variety of phenomenology abandons claims of first philosophy, if it seeks to validate its first-person tools and to combine them with suitable second-person and third-person tools, if it acknowledges the relevance of scientific data and theories for itself and the amenability of its subject matter to empirical sciences, then we may call this version of phenomenology naturalized. Note that this suggestion does not imply at all that phenomenology as first philosophy is useless, as I will immediately show.

Keeping in mind that 'the naturalization of phenomenology' embraces a whole range of various versions and nuances of mergers of commitments and methods, I plead for a 'weak' (i.e., partial or limited) naturalization of phenomenology for different reasons. (1) The ambitious aim of phenomenology as a rigorous science has not been achieved; so far phenomenology has not provided us with a basis from which all further claims to knowledge, objective science included, can be scrutinized and epistemically justified once and for all. (2) Despite this moderation, transcendental phenomenology has proven to be a viable analytic tool in the study of human cognition. Other disciplines have started to profit from analyses provided by transcendental phenomenologists. (3) As I have argued above, even classical ('domestic') phenomenology is not limited to first-person accounts of isolated conscious experiences. Phenomenology aims at mutual assessment and validation by

peers and is open to interdisciplinary or mutual enlightenment. It should be informed by the best available scientific data and theories. (4) Finally, naturalized phenomenology might be in danger of becoming oblivious to the implicit assumptions of science, which cannot be reflected upon within the framework of science itself. Similarly, there are implicit assumptions even of pure phenomenology that cannot be aptly uncovered within the transcendental attitude, which is the reason why Husserl postulated a phenomenological critique of the transcendental naiveté (cf. Husserl 1967).

Thus, even if one thinks that the idea of a rigorous transcendental phenomenology is not justified in the last respect, particularly concerning Husserl's claim that the objective sciences are in need of a kind of explanation or understanding that only transcendental phenomenology can achieve (cf. Husserl 1970, §55), we should nonetheless consider transcendental phenomenology in the expanding interdisciplinary discourse. The transcendental voice may save current dialogues, including naturalized phenomenological approaches, from the (not inescapable) naïveté of strictly objectivistic approaches and from the dangers of a reductionism or eliminativism in phenomenological disguise. To be informed, enlightened by empirical disciplines is one thing; to keep being informed by transcendental reflection is the flipside of a 'mutually enlightened' (Gallagher 1997) discourse.

Put differently, what I am proposing amounts to a division of labor among phenomenologists: let some rub shoulders with 'objectivist', working scientists while others hold the fort of rigorous, transcendental philosophy. This seems, to me, the most fruitful way of doing phenomenology. This is in close agreement with the account of the relevance of phenomenology to contemporary cognitive science and consciousness studies, which Gallagher and Zahavi (2008) have offered lately. Pointing out the importance of an exchange between phenomenology and the empirical disciplines they equally remind us of the significance of the transcendental attitude. 'There is no incoherence in claiming that phenomenology should be as informed by the best available scientific knowledge, while at the same time insisting that the ultimate transcendental philosophical concerns of phenomenology differ from those of positive science' (Gallagher and Zahavi 2008, 221).

Not all phenomenologists will agree with my conciliatory proposal. Just to name two recent authors, Simon Glendinning and Dermot Moran regard phenomenology as inseparable from an anti-naturalistic attitude and thus as *the* alternative to naturalism in philosophy of mind and in Western modern thought in general. Yet, as David Hume once put a similarly intricate issue, '[r]easonable men may be allowed to differ, where no one can reasonably be positive' (Hume 1779, 30). We cannot reasonably be positive about the ultimate relationship between transcendental phenomenology and the empirical sciences, and hence we differ. It's up to us to turn disagreement into collaboration.[11]

References

Damasio H (2001). Words and Concepts in the Brain. In: Branquinho J (ed) The foundations of cognitive science. Clarendon, Oxford, 109–120

Dennett D (1991) Consciousness explained. Little, Brown, Boston/New York

Drummond J (1990) Husserlian intentionality and non-foundational realism. Kluwer, Dodrecht

Embree L (2007) Reflective analysis. Zeta Books (eBook, and print on demand)

Gallagher S (1997) Mutual enlightenment: Recent phenomenology in cognitive science. J Conscious Stud 4(3):195–214

Gallagher S (2003) Phenomenology and experimental design: toward a phenomenologically enlightened experimental science. J Conscious Stud 10(9–10):85–99

Gallagher S, Sørensen JB (2006) Experimenting with phenomenology. Conscious Cogn 15:119–134

Gallagher S, Zahavi D (2008) The phenomenological mind. An introduction to philosophy of mind and cognitive science. Routledge, London/New York

Glendinning S (2007) In the name of phenomenology. Routledge, London/New York

Heidegger M (1962) Being and time (trans: Macquarrie J, Robinson E). Harper & Row, New York

Heidegger M (1985) History of the concept of time (trans: Kisiel Th.). Indiana University Press, Bloomington, IN

Hume D (1779/1993) In: Gaskin JCA (ed) Dialogues concerning natural religion; the natural history of religion. Oxford University Press, Oxford

Husserl E (1952) Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Drittes Buch: Die Phänomenologie und die Fundamente der Wissenschaften. Nijhoff (Husserliana; V) The Hague

Husserl E (1959) Erste Philosophie (1923/24). In: Boehm (ed) Zweiter Teil: Theorie der phänomenologischen Reduktion. Nijhoff. (Husserliana; VIII), The Hague

Husserl E (1964) Philosophy as a rigorous science. In: Lauer Q (ed) Edmund Husserl. Phenomenology and the crisis of philosophy. Harper & Row, New York

Husserl E (1967) Cartesian meditations (trans: Cairns D). Nijhoff, The Hague

Husserl E (1970) The crisis of European sciences and transcendental phenomenology: an introduction to phenomenological philosophy (trans: with an Intro. by Carr D) Northwestern UP, Evanston, IL

Husserl E (1973) Experience and judgment: investigations in a genealogy of logic (trans: Churchill JS, Ameriks K). Northwestern UP, Evanston, IL

Husserl E (1982) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. First book: general introduction to a pure phenomenology (trans: Kersten F). Kluwer, Dordrecht/Boston/London

Husserl E (1988) Analysen zur passiven Synthesis. Aus Vorlesungs- und Forschungsmanuskripten (1918-1926). (ed. Fleischer) Nijhoff (Husserliana; XI) The Hague

Husserl E (2001) Logical Investigations. 2 vols (trans: Findlay JN, pref: Dummett M, ed: with new intro. by Moran D) Routledge, London/New York

Ihde D (1977) Experimental phenomenology: an introduction. Putnam, New York

Marbach E (1993) Mental representation and consciousness: towards a phenomenological theory of representation and reference. Kluwer, Dordrecht

Merleau-Ponty M (2002) Phenomenology of perception (trans: Smith C). Routledge, London

Merleau-Ponty M (1963) The structure of behavior (trans: Fisher AL, foreword: Wild J). Beacon, Boston

Merleau-Ponty M (1964) Signs (trans: with an intro. by McCleary RC). Northwestern University Press, Evanston

Moran D (2000) Introduction to phenomenology. Routledge, London/New York

Moran D, Mooney T (eds) (2002) The phenomenology reader. Routledge, London/New York

Petitmengin C (2006) Describing one's subjective experience in the second person: an interview method for the science of consciousness. Phenomenol Cogn Sci 5(3–4):229–269

Roy J-M, Petitot J, Pauchoud B, Varela FJ (1999) Beyond the gap: An introduction to naturalizing phenomenology. In: Petitot J, Varela FJ, Pachoud B, Roy J-M (eds) Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science. Stanford University Press, Stanford, 1–80

Schmicking DA (2003) Hören und Klang: empirisch phänomenologische Untersuchungen. Königshausen und Neumann, Würzburg

Schmicking DA (2005) Is there imaginary loudness? Reconsidering phenomenological method. Phenomenol Cogn Sci 4(2):169–182

Sokolowski R (2000) Introduction to phenomenology. Cambridge University Press, Cambridge

Spiegelberg H (1975) Doing Phenonemology: Essays on and in Phenomenology. Nijhoff, The Hague (Phaenomenologica; 63)

Spiegelberg H, with collaboration of Schuhmann K (1994) The phenomenological movement: a historical introduction. 3rd rev. and enlarged edn. Kluwer, Dordrecht/Boston/London

Thompson E (2007) Mind in life: biology, phenomenology, and the sciences of mind. Harvard University Press, Cambridge/London

Varela FJ (1996) Neurophenomenology. A methodological remedy for the hard problem. J Conscious Stud 3(4):330–349

Varela FJ, Shear J (1999) First-person methodologies: what, why, how? J Conscious Stud 6(2–3):1–14

Vermersch P (1999) Introspection as practice. J Conscious Stud 6(2–3):17–42

Wrathall MA (2006) Existential phenomenology. In: Dreyfus HL, Wrathall MA (eds) A companion to phenomenology and existentialism. Blackwell, Malden, MA/Oxford, 31–47

Zahavi D (2003) Husserl's phenomenology. Stanford University Press, Stanford, CA

Towards a Formalism for Expressing Structures of Consciousness

Eduard Marbach

The objective of developing a formalism in Phenomenology is threefold. First, a formalism should help the phenomenologist to put down in more precise and stable form what this or that means in his or her attempt reflectively to describe conscious experiences. Second, it should facilitate communication of phenomenological findings. Third, it should advance the elaboration of agreement procedures among researchers of consciousness using first-person methodologies. To enable intersubjective agreement concerning essentially subjective, first-person data would, in itself, seem to be a valuable goal for an investigation of consciousness within philosophical Phenomenology.[1] Moreover, such agreement is no doubt requisite for successfully integrating phenomenological data within scientific studies of consciousness in the Cognitive Sciences.[2]

The following chapter first presents a brief background account relevant to the present proposal. This account includes a few remarks about the problem of scepticism concerning phenomenological data which would seem to undermine claims that such data can be of scientific value. It then recalls Husserl's view of Phenomenology as a rigorous science and comments on the fact that Husserl himself

E. Marbach
Prof. emer. for Phenomenology and Philosophy of Mind, Institute of Philosophy,
University of Bern, Switzerland

[1] Daniel Schmicking (2005), for example, mentions Marbach's (1993) "project of developing a phenomenological notation" and remarks: "So there are already glimpses of the appropriate means to develop a phenomenology that would be comparably more precise in articulating its outcomes and more amenable to intersubjective verifiability". He then suggests that "phenomenologists should set themselves the task of developing agreement procedures, of testing the practicability of such procedures, and of developing a common phenomenological notation that enables authors to 'transcribe' … structures of consciousness" (p. 180f.).

[2] See, e.g., David Chalmers' (1999) statement: "In my opinion, the development of more sophisticated methodologies for investigating first-person data and of formalisms for expressing them is the greatest challenge now facing a science of consciousness. Only by developing such methodologies and formalisms will we be able to collect and express first-person data in such a way that it is on a par with third-person data, so that we can find truly systematic and detailed connections between the two" (p. 10).

only very rarely had recourse to formalisms in the context of reflective descriptions of consciousness. The contribution then proceeds in two main parts; Part I: Towards a Formalism for Philosophical Phenomenology, Part II: An Application to Scientific Studies of Consciousness.

Scepticism with regard to the validity of phenomenological results has often been expressed. This is true with respect to the philosophical movement of Phenomenology in the Brentano-Husserl tradition as much as with respect to contemporary investigations of conscious experience that appeal to first-person methods and often use the terms 'phenomenology' and 'phenomenological' quite independently of that tradition to designate aspects of experiences, their so-called "phenomenal", or "subjective", or "qualitative" character, or their ways of seeming or feeling for someone, etc., the description of which, and even the question of their very existence, give rise to disputes. To illustrate such scepticism, consider, for example, Thomas Metzinger (1995): "The phenomenological movement was the first comprehensive attempt to establish the conceptual and epistemological basis for a systematic, autonomous *science of consciousness*. That is its merit. Its failure was mainly due to its reliance on the evidence of inner perception. This proved to be an untenable approach of data collection because it lacks a reliable procedure for eliminating false observations and statistical inconsistencies, especially in the case of conflicting statements" (p. 25).[3] For many years, as is well known, Daniel Dennett (e.g. Dennett 1991) also forcefully attacked "the Fantasy of First Person Science" and criticized what, with particular reference to the tradition of Husserlian Phenomenology, he calls "auto-phenomenology", which is based, as Dennett views it, on notoriously unreliable introspection.[4] Recently, Alva Noë (2007) observed: "As it happens, I am persuaded by the sorts of phenomenological claims I am about to put forward … But if the distinct phenomenological claims I make are wrong, then that fact can also, in itself, help to demonstrate that phenomenology is a domain for substantive dispute" (p. 235). And Uriah Kriegel (2007), after discussing a number of disputes about what is phenomenologically manifest in conscious experience, puts it thus: "This multiplicity of phenomenological disputes is in one way exciting, but in another unsettling. It may be especially dispiriting if we have no firm handle on the disputes and how to adjudicate them" (p. 120f.). Similar remarks from a variety of writers could be cited.

As is well known, Husserl himself thought of his phenomenology as a rigorous science. As late as 1919, he wrote in a letter: "It is, I like to say, the pride of transcendental phenomenology and its hallmark as a rigorous science that false sentences can occur in it, sentences false in the strict sense of logic which always prove themselves to be such by and against truths".[5] Given that Husserl was formed as a

---

[3] For a critical comment on Metzinger's view, see Zahavi (2007), p. 27f.

[4] For an instructive detailed discussion of the problem of scepticism with regard to phenomenological claims, with special attention to Dennett, see Roy (2007).

[5] See "Edmund Husserl: A letter to Arnold Metzger" (September 4th, 1919), introduced and translated by Erazim V. Kohàk, *The Philosophical Forum*, 21 (1963–1964): 48–68; the quoted passage is from p. 64f.

mathematician and logician, it is, however, somewhat surprising to realize that he himself very rarely made use of a notation in his analyses of conscious experiences. One of the few places where Husserl did apply symbols to the phenomenology of conscious experiences or mental acts occurs in "§39. Presentation in the sense of the objectifying act and its qualitative modification" in the Fifth of the *Logical Investigations* of 1901. Explaining the radical difference between the qualitative modification (what Husserl later calls 'neutrality modification'[6]) of a positing act into an act that "merely presents" its object on the one hand and, on the other, the wholly different operation of generating a presenting objectivation relative to a mental act, Husserl introduces the following symbols:

$$O, V(O), V\left[V(O)\right]...$$

He comments that O designates some object, $V(O)$ the presentation of O, and he points out explicitly that the operation of the presenting objectivation be able to be "*in infinitum* iterated", whereas the qualitative modification is a non-recurring operation. As an example, Husserl mentions a house to which corresponds a presentation of the house, and to the presentation corresponds a presentation of the presentation, to the judgment the presentation of the judgment, etc. On the other hand, once the "belief" of a positing mental act is modified into a "mere presenting", we can at best revert to the belief, but there is no modification recurring and continuing in the same sense.[7]

In the context of the present contribution, the most important place where Husserl did use symbols is to be found in a research manuscript from around 1911/1912.[8] In this text, he sets out to clarify questions pertaining to the phenomenology of mental acts of representation (*Akte der Vergegenwärtigung*) as opposed to acts of simple perception. In the course of discussing questions of internal consciousness, internal reflection, and the concept of reproduction, Husserl notes: "The external perception is perception. And if the modification of the perception is then a corresponding memory, we have the remarkable circumstance that the corresponding memory is not only memory of the perception but that the modification of the perception is also memory of what was perceived. I must make this clearer" (p. 370). Husserl then gradually turns to a few symbols in order to represent his reflective findings relative to the relationship of representation, reproduction and perception. He emphasizes that the term 'reproduction' is meant to serve for designating the "representation of *internal consciousness*" and that it should not falsely be taken to mean "that the originary experience belonging to the earlier internal consciousness is now being

---

[6] See *Ideas I,* 1913, §112.

[7] See also the excellent discussion of Husserl's view of phenomenology and the application of mathematical concepts to phenomenology by Jeffrey Yoshimi (2007); with regard to the example just presented, see p. 274. Yoshimi provides a couple more examples where Husserl deploys mathematical methods in phenomenology (see pp. 274–277, in particular).

[8] See Edmund Husserl (2005), Text No.14 (1911–1912): 363–377 (Translation from *Husserliana* 23, The Hague: M. Nijhoff 1980, Text Nr. 14: 301–312).

produced again, as if an echo, reflection, afterimage of the earlier internal consciousness, although weak, were coming back". "In truth, it is a representation; and representation is a new kind of *act*" which is "separate" from "the representation of a physical event"; the latter "must not be called reproduction". "The natural event is not produced once again. It is remembered; it stands before consciousness in the manner of what is represented" (p. 372). He then notes: "Now let us consider the remarkable relationship between the two representations to be compared here – representations that obviously differ from one another intrinsically (*an sich*)" (p. 372). Using the following symbols from the English translation: 'A' for an act, 'a' for an external object, 'P_e' for an external perception, 'R_a' and 'R_i' for an external or internal representation, respectively, and 'Rep' for a reproduction, we get:

1. Over against $A \equiv P_e$ there stands R_i or, as we can now write, Rep (P_e), the reproduction of the external perception.
2. Over against P_e, understood as $P(a)$, the perception of the external object a, there stands R_e; that is to say, $R(a)$, the representation of a (p. 372).

Applied to the perception of a house, Husserl's question is, how the reproduction of the perception of a house compares to the representation of the house, to which he answers that, while excluding acts of meaning and differences in attention, "surely we must say that the phenomena in both cases are the same, that an essential law obtains here according to which

$$\text{Rep}\left(P_e\right) = R_e$$

and this is valid for every perception that is not internal consciousness (although for the latter too, of course, as tautological limit case)." (p. 373).[9]

In spite of such examples of Husserl's use of symbols, the fact that formal notations in the context of his many analyses of conscious experiences and their intentional correlates are almost totally absent may also serve as a warning not to expect too much of them. Husserl's insistence on direct seeing and describing the essential structure of what is given in reflection upon conscious experiences and their intentional correlates is well known. It may also be worth recalling here Husserl's understanding of the 'essence' or 'essential structure or form' of experiences which, in a mathematical spirit, has to do with defining the conditions of "the possibility of experience" while being unconcerned with empirical matters of fact. As he put it in 1907: "The essence of experience, which is what is investigated in the phenomenological analysis of experience, is the same as the possibility of experience, and everything established about the essence, about the possibility of experience, is *eo ipso* a condition of the possibility of experience".[10]

---

[9] In Andrea Borsato (2009), Innere Wahrnehmung und innere Vergegenwärtigung, there is a fine detailed discussion of these and related issues, giving particular attention to a comparison between Brentano's and Husserl's views that were modified substantially over the years.

[10] See Edmund Husserl (1997), §40, p. 119.

It is obviously crucial for phenomenological descriptions of the essences claimed to be the very conditions of the possibility of the conscious mental activities that the descriptions be based on *actual reflection* upon the activities in question and their correlates. As Husserl has it: "the phenomenological method operates exclusively in acts of reflection" (*Ideas I*, §77). Reflective analysis of the details of conscious experiences, according to Husserl, has to be carried out with the help of *descriptive concepts* which express the essential nature of these experiences by drawing "immediately from simple intuition" (*schlichte Anschauung*), unlike, e.g., geometrical concepts which are "'ideal' concepts, expressing something which cannot be 'seen'".[11] Even though "deductive theorizings are excluded from phenomenology" (§75), Husserl concedes that "*mediate inferences* are not exactly denied to it". However, stressing again that all phenomenological cognitions concerning conscious experiences "ought to be descriptive, purely adjusted to the sphere of immanence", he explains that "inferences, non-intuitive modes of procedure of any kind, only have the methodical function of leading us to the matters (*den Sachen*) which a subsequent direct seeing of essences must bring to givenness" (ibid.). Prior to actual intuition, analogies may obtrude themselves, say, among the structures of conscious experiences of various kinds, suggesting conjectures about essential connections from which further inferences may be drawn. But, as Husserl emphasizes, "ultimately an actual seeing (*wirkliches Schauen*)" of the connections "must redeem the conjectures. As long as that has not occurred, we have no phenomenological results" (ibid.). In the present context, it is, however, particularly intriguing to keep in mind what Husserl adds immediately after: "The question obtruding itself of whether in the eidetic sphere of the reduced phenomena … there could not, besides the descriptive procedure, be an idealizing procedure that substitutes pure and strict ideals for the intuitive data, <ideals> which then might even serve as the fundamental means for a *mathesis* of experiences – as counterpart of the *descriptive* phenomenology – is thereby of course not answered" (§75, p. 169).

Towards a Formalism for Philosophical Phenomenology

In Marbach (1993), some inspiration for wanting to introduce a notation in view of improving the descriptive phenomenological method came from Frege. The notation should serve the purpose of expressing with precision *reflectively* elaborated forms of consciousness following methodological devices of Husserlian phenomenology.[12] Indeed, Frege characterized his *Begriffsschrift. Eine der arithmetischen*

---

[11] See Edmund Husserl (1913/1983), §74.

[12] The formalism of the "phenomenological notation" in its present development follows still closely Husserl's methodological constraints. In a research project currently funded by the Swiss National Science Foundation, a more advanced formalism with special attention to its syntax should be forthcoming thanks to the collaboration of Sebastian Leugger, in particular.

nachgebildete Formelsprache des reinen Denkens from 1879 in ways which invited an attempt to envisage a script that would prove to be suitable, *mutatis mutandis*, for expressing the conceptual content (begrifflichen Inhalt) revealed by phenomenological analysis of the contents of conscious experiences. It seemed particularly encouraging to find Frege saying the following at the end of his short paper "Über die wissenschaftliche Berechtigung einer Begriffsschrift" (Frege 1882b):

> … at any rate an intuitive representation of forms of thought has an importance reaching beyond mathematics. May therefore philosophers, too, pay some attention to the matter! (p. [56], 114)

Despite the different goals of Frege's new script, modelled on the language of arithmetic, it seemed particularly relevant to aim, with Frege (1882a), at

> expressing a content by means of written signs in a way that is more precise and easier to survey than it would be possible by means of words. (p. [1], 97)

The envisaged "phenomenological notation", like Frege's Begriffsschrift, should also help to avoid misunderstandings by others and at the same time mistakes in one's own thinking (Frege 1882b, p. [48], 106).

No doubt, one of Husserl's most productive achievements consists in his detailed intentional analyses by which he was able to make concretely intelligible that the very ways of intentionally referring to something are simultaneously so many modes of inner consciousness. Put another way, the mature Husserlian method of elucidating mental phenomena as *consciousness of something* is at the same time *objectively* (noematically) *and subjectively* (noetically) oriented.[13] The analysis of the intentionality of mental phenomena qua conscious experiences is concerned with noetico–noematic correlations. In regard to the objective side, experiences are conscious *of something*; in regard to the subjective side, the experiences themselves are at the very same time *conscious* of something. To be *conscious of something* in various ways is a matter of the very performance of mental acts. Thus, differences of consciousness are decisive for the ways of achieving intentional reference to something. For example, mental acts of perceiving something present and of imagining something non-present are distinctly different from one another precisely as momentary experiences subjectively lived through by a person. By the same token, they achieve distinctly different intentional references, or object-relations, in virtue of their ways of presenting something objective. From early on, Husserl himself pursued the analysis of conscious experiences in terms of intentionality as a "descriptive science … of the *multiple forms of consciousness as consciousness of something*, with all its constituents which can be distinguished in internal intuition".[14]

Within the scope of this contribution, the focus will be on pre-linguistic mental acts of intuitive representation (*anschauliche Vergegenwärtigung*), such as remembering, imagining, picturing something, and combinations of such acts. These acts are analyzed in terms of modifications of the basic form of perceptual consciousness

---

[13] See *Ideas I*, 1913/1983, especially Part III, chapters 3 and 4.

[14] See Husserl (1977), p. 23.

with its intentional reference to something in its presence. Put differently, they are analyzed in terms of intentional relations between acts of presenting something and of representing something. And these intentional relations are described in terms of *intentional implication and modification*. The basic idea of the phenomenological notion of *intentional implication* is one of "being contained" within a compound (*intentionales Beschlossensein*). For the reflective analysis of the phenomenological forms it is then a matter of unfolding that which is contained, implicated, within the whole mental act. What is said to be "intentionally implied" concerns the constituent, hence non-independent, components, parts or moments of a mental act as a whole. The whole mental act *as a unity of consciousness* can be said actually to consist in the multiplicity of its parts or moments. It should become clear below that the analysis of intentionality in terms of intentional implication is at the same time analysis of intentionality in terms of intentional modification.

Now, in connection with the task of defining such forms of consciousness, the *phenomenological notation* should prove to be illuminating especially in the following two respects. First, the *non-phonetic signs* of the notation may help to make the descriptive language of phenomenology more concise and transparent regarding those forms as they are given in reflection. If the forms of representation are to be defined as so many unities, or unified structures, of the performance of mental acts containing a multiplicity of analytically separable moments within themselves,[15] then the artificial signs of the notation may enable the phenomenologist clearly to designate the particular components of a mental act that are reflectively separated from one another and simultaneously make their phenomenological unity more readily graspable than the all too longwinded expressions from ordinary languages would permit. Second, the notation should contribute to a better understanding of the varieties of the so-called intentional correlation between the objects (things, events, etc.) referred to and the mental acts themselves, considered as forms of consciousness of something.

In short, if it is the function of consciousness, of intentionality, to establish (intentional) reference to something objective in one way or another, then the phenomenological notation can provide precise expressions for the *varieties of intentionality*. At the same time, more precision and distinctness can be obtained regarding the property of "*consciousness*", or of the experiential "what-it-is-likeness". It should once more be emphasized, though, that the proposed formalism for Phenomenology relates to *reflectively* given entities only. The notation is meant to improve the reflectively-oriented fundamentally descriptive theoretical language of Phenomenology, and not our ordinary ways of talking about conscious experiences. Using the notation especially for the more complex acts of intentionally referring to something helps to formulate the findings of the easily divagating reflective attitude in a more condensed manner, and this is valuable because it shows how the mental acts are so many more or less complex *unities of performance* exhibiting

---

[15] For an extensive methodological presentation of this analytical approach to conscious experiences, see Kern (1975), "III. Abschnitt. Philosophische Methodenlehre", in particular "1. Kapitel: Die reine Reflexion" and "2. Kapitel: Wesenserkenntnis".

similarities and differences among themselves. As Frege (1882b) pertinently put it with regard to his *Begriffsschrift*:

> The script provides the possibility of *keeping a lot present at the same time*, and even if we can only fix our eyes upon a small part of it at any given moment, we nonetheless retain a general impression also from the rest, which is, whenever we need it, at once at our disposal. (p. [53], 111, emphasis EM)

For present purposes, the potential of the phenomenological notation along these lines may be illustrated with a few formulae using the following signs:

– Triplets of upper case letters serve to designate mental activities; thus 'PER' for the activity of 'perceiving' which, in given cases may be visual, auditory, tactual, gustatory, olfactory or any multimodal combination thereof; 'REM' for 'remembering', 'IMA' for 'imagining', 'PIC' for 'picturing' or 'image consciousness', 'PRE' for 'presenting', 'REP' for 'representing'.

– Lower case letters ('x', 'y', 's') designate intentional objects, i.e. objects reflectively considered as correlates of mental activities. In the present context, the value range of 'x' is any individual, spatio-temporally located object, situation, event, considered as correlate of a mental activity; 'y' serves to designate the physical picture, 's' or 's`' designates one's present or represented surroundings. On reflection, it will become clear that x's and y's are parts of s's singled out in one way or another by activities of PRE or REP.

– A pair of curly brackets, '{…}', serves to designate the very *theme* of a reflective analysis. Thus, expressions like {REM x}, {IMA IMA x}, {REM PIC x}, etc. are meant to indicate that the mental acts mentioned inside the brackets are to be submitted to reflective analysis in order to determine their phenomenological forms.

– A pair of parentheses, '(… …)' together with 'x' etc., is used to designate the reflective finding of the *intentional correlation* between an *actually* occurring mental act as a whole and its object(s). Thus expressions such as '(PER)x', '(IMA)x', etc. designate the fact that an actually occurring perception, imagination, etc. has x as its intentional object or correlate. These act-wholes are to be reflectively analyzed in terms of phenomenological forms or structures of presenting and representing the intentional object. Putting the '(…)'-parentheses around the expressions for the multiple components of the *act*, rather than around the expression for the intentional correlate, is precisely meant intuitively to capture the *unity of an act as a whole*.

– Pairs of square brackets, '[…]', surrounding expressions for mental activities and put *inside* the expressions for the act as a whole surrounded by parentheses, designate the reflective finding of an occurrence of an *intentional implication* or *modification* of a mental activity contained within another activity. Crucially, the expression of an activity surrounded by such brackets is meant to be indicative of the fact that such an intentionally implied activity is involved in the *mode of non-actuality*; it is no longer actually performed but is, precisely, experienced as being only represented (*vergegenwärtigt* in Husserl's sense of *reproduction* – see above) in one way or another in one's actually representationally (vergegenwärtigend) referring to an object. Thus, the expression '(REP[PER])x' designates the

reflective finding that the actually occurring representational act in its intentionally referring to x implies within itself a non-actual activity of perceiving x. Forms of representational experiences, (REP)x, relevant in the present context are forms of IMA, REM, PIC, and combinations thereof; they all intentionally imply an activity of perceiving, or more than one, in ways further to be specified.

– The correlates of the act and its components are written *outside* the parentheses and brackets, simply as x, y, s. It is true, however, that the expressions for the objects could also be put within parentheses or brackets of their own, possibly doing better justice to Husserl's method of the transcendental phenomenological reduction according to which objects are sometimes said to be "put within parentheses", thus in the notation as (x), (y) and (s) next to expressions surrounded by '(…)' and '(…)', respectively.

– A horizontal stroke, '_____', called "foundation-stroke", serves to account for the reflective finding that all representational experiences involve a simultaneously occurring presentational experience on the basis of, and in contrast to, which one's intentionally referring to something represented takes place; the compound expression is set in subscript position: $\overline{(PRE)s}$

– In addition, this stroke also serves in cases where a representational experience involves a simultaneously physically present or represented carrier or foundation for the representational function, such as in cases of picturing something. Thus, an expression such as

$$\frac{(REP - [\overline{PER}]) \; x \mid x}{\overline{(PER)} \; y}$$

– designates part of the form of the mental activity of pictorially representing some x that is grounded in a simultaneous perceiving of the picture (carrier) y, in which the x appears. The vertical stroke '|' between 'x' and 'x over y' is used to capture the finding that, with (PIC)x, intentional reference is made to a peculiar "double object" (i.e., the depicted real or fictional x just in so far as it appears in the physical picture y as pictorial object 'x over y').

– A sign of the form '⊢', called "belief-stroke", when put in front of a pair of brackets serves for expressing the fact that a given represented mental activity is experienced with the force of "belief" (or "positionality"), '⊢ (…)', and, when put in front of the symbol for the intentional correlate, that this is taken for something in the real world.

– On the other hand, a sign of the form '–', called "neutrality-stroke", when put in front of a pair of brackets and/or in front of the symbol for the intentional correlate, serves for indicating a conscious operation of neutralizing, i.e. of suspending one's belief, either with regard to a represented activity or to an intentional object.

– Written in subscript position, the letter 'p' serves as a temporal index for 'past', and where necessary, the letter 'i' will be written at the very beginning of a formula to designate the I-consciousness that is involved in representational experiences.

Given the remarkable neglect of subjectively-oriented analyses in many quarters, often in the name of the thesis of the so-called "transparency of experiences", according to which subjects of experiences manage *only* to become aware of the properties of what these are experiences *of*, the notation may help draw attention in concise ways to subjective (noetic) components of conscious experiences. These (noetic) components do not coincide with the representational (noematic) content but are instead instantiated correlatively with the representational (noematic) content.[16]

Consider first the following two examples of acts of representational consciousness to be reflectively analyzed: (1) Remembering object x; (2) viewing a picture, or having image consciousness, of object x. Each time, ask yourself, how are the objects x, y, s, etc. given in each one of the experiences of intentionally referring to them? Alternatively, ask yourself, what is it that you do in order to have x, y, s given to you in such and such a way? Proceeding in this way reflectively, so to speak *backwards* from the intended objects x, y, s out there (e.g., in the present or past real world, in a pictorial world, or in some fictional setting) to the modes of being given of these objects in one's conscious experiences, one can determine stepwise structurally clearly distinct formulae for the phenomenological forms of these experiences.

(1) {REM x}, what is it, to remember object x, e.g., Lake Como as seen from the aeroplane?

In a very first approximation, the activity of remembering a view of the Lake is an activity of referring to an object in the real world *in its absence, believed to have been presently given* in the *past*. More precisely, it is an intentional referring to an object in its *not actually being present* in the sense of appearing here and now, relative to the presently occupied perceptual point of view. However, reflecting upon the way of being given of the Lake in an actually performed act of remembering, which is a form of representational consciousness, reference to the Lake is established in such a way that it is as if the Lake were given to me again in perception, i.e. in a spatial orientation to a perceptual point of view of mine. The phrase "as if" (or "as it were" or "quasi" for that matter) is, however, indicative of some *modification* of one's activity of perceiving, to the effect that one's original perceptual referring to something in its present actuality is altered as regards the mode of consciousness: it is involved as perceiving in the mode of non-actuality. A fundamental condition of the possibility of rememberingly representing an object (thing, event, etc.), then, is this: To be now actually intentionally referring to a represented x is, at the very same time, to be representing (in Husserl's sense of reproducing – see above) a mental activity of presentation (in the mode of non-actuality) relative to which x was given to me as present. It appears, then, that somehow a plurality of

---

[16] As the notation was introduced in Marbach (1993), emphasis had really been given to the subjectively-oriented, noetic side of experiences, at the cost of the objectively-oriented, noematic analysis. There is clearly a requirement to supplement the notation in this regard; presently, however, only minor amendments will be taken account of. The current research project mentioned above will offer a more developed version regarding the noematic side.

mental activities is at work when one is establishing reference to the *one* intended represented object, say, Lake Como. This much may be put as follows, using the signs introduced above:

(1a) $$(REP\ p \vdash [PER])\vdash x$$

This means: actually representing x, believed to be real, by means of representing (reproducing) a perceiving of x believed to have occurred in the past.

A closer analysis reveals the following essential connection: Reflecting upon the way one is given Lake Como in the example, it can be pointed out that to be intentionally referring to something absent can only make sense if, *at the very same time*, one is also referring to something else that is presently given. To say that something is absent, represented, is not simply to say that it is not present. Rather, it is also saying that, at the very same time, something else is present. The Lake in its not being present can appear as it were in correlation with one's represented perceptual point of view (looking out of the window up in the aeroplane) only if, at the very same time, something other than the Lake referred to does actually appear in correlation with one's actually occupied point of view. Thus, there is a further condition of the possibility for something represented to be given as such, namely the fact of the necessary *dependency* of activities of representing upon (or their relativity to) some simultaneous actually performed presentation. This reflective finding can be captured thus:

(1b) $$\frac{\qquad\qquad (REP\ p \vdash [PER])\vdash x}{(PRE)\ s}$$

This means: while grounded in actually presenting some surroundings s, actually representing x, believed to be real, by means of representing (reproducing) a perceiving of x believed to have occurred in the past.

A further condition of the possibility of intentionally referring representingly to the remembered object appears to be this. All the intentional components analyzed so far appear to be united in the subject that performs the activity of representing something. A mental activity of representing x does *not* consist of multiple fragmented intentional components. Quite to the contrary, a mental activity of representing something, say, rememberingly, is experienced as a unity of consciousness, a unity of performance, in virtue of which reference to something identical and objective, the x, is established. This complexity of the components of a representing activity is simultaneously unified in an I-consciousness. The unification of temporally and spatially discontinuous moments takes place in an I-consciousness: being more or less lively or attentively conscious of representing x, i.e. of being as it were "there and then" (sitting at the window in the aeroplane), while simultaneously, albeit less lively, being conscious of presenting s, thus of being "here and now" (at my desk). The phenomenological claim is that such I-consciousness is identical neither with one's consciousness of one's actual bodily position in the here and now alone, nor with one's consciousness of any of the represented bodily quasi-positions

by themselves. Much rather, I-consciousness in acts of mentally representing x can be said to occur as a subjective unifying reference of the plurality of the actual moments and of the non-actual ones that are intentionally implied in the unity of performing one and the same activity converging objectively on one and the same x. As Husserl once put it:

> An act is as it were an intending (in a very broad sense) … all these <acts> are act-connections which, even in their often overwhelming implications, make up the unity of one act and bring to consciousness an objective correlate which thereby stands "over and against" the I.[17]

In sum, then, the phenomenological form underlying an activity of remembering x may be put like this:

(1c) $$\quad\quad\quad\quad\quad\quad\quad i \underline{\quad\quad\quad} (REP\ p \vdash [PER]\ ) \vdash x$$
$$\quad\quad\quad\quad\quad\quad\quad (PRE)\ s$$

This means: I, while grounded in actually presenting my surroundings s, am actually representing x, believed to be real, by means of representing (reproducing) a perceiving of x believed to have occurred in the past.

Notice that while one is performing a representational act, e.g. of remembering x, it is the x (the past event, etc.) that one is actually objectively referring to. The reproduced activity, e.g. of perceiving, said to be intentionally implied in the representational act, is *not* itself objectively given in one's performing the representational act. It comes only to the fore as an object in the phenomenological description based on reflection upon the act-components *by means of which* the intentional reference to something represented is established.

Turning now to a representational act of image consciousness or of picturing, some of the reflective findings already discussed do recur, as formula (2a) of the notation corresponding to the phenomenological form of the act will show. Here, special attention will be paid to the peculiarity of the "double object" intentionally referred to in one's image consciousness as reflective phenomenological analysis shows in answering the question of how an object that is depicted in a picture, or how a pictorially appearing object, is given.

(2) {PIC x}, what is it, to have an image consciousness, or to view a picture, of object x?

Representational acts of PIC x can in a first approximation be seen as a mode of referring to x in its absence such that something of the way of x's appearing is literally preserved in something else. For, instead of being merely referred to by means of a mentally represented point of view – as, for example, in REM x – the x referred to by means of a pictorial representation does, somehow, appear as it were in a stable physical picture or in a statue, designated by 'y'. It would seem, then, that in pictorial representations something real and appearing permanently in

---

[17] See Edmund Husserl (1989), Appendix XII (1917), p. 344f – See also Kern (1975), §21. Das "Subjekt" des Verstandes: das reine Ich, p. 65ff.

correlation with an actually occupied perceptual point of view stands for something else, something absent, that tends to "disappear" when it is represented purely mentally, as in REM x or IMA x. With PIC x, unlike what happens with regard to the x referred to in the simpler forms of REP x, PER is not only implied in the mode of non-actuality. Rather, some PER is *actually* involved in one's referring to x. The present object y is actually perceived in order to establish intentional reference to x in its absence. When PIC x is performed, x and y are, however, not simply two different entities outside one another. Instead, one is taken to be a representation (*Darstellung*) of the other, y is taken to be a picture *showing* x. *Something* of x at least, namely something that would be perceivable if x itself were present, must therefore also be perceivable in y. Oddly, then, it would seem that y is not just y, nor is it entirely x. Pictures are paradoxical entities bestowed with a kind of *double* or *dual reality*, as has also been pointed out in cognitive psychological literature.[18]

Reflective analysis of the phenomenological form of image consciousness attempts to elaborate the very conditions of the possibility of pictorial representations such as pictures, sculptures etc. It tries to make the double or dual status of pictures intelligible by pointing out a *twofold intentionality* in the unity of one's performing an act of referring to x by means of a picture. To take an example, if the object x, say, a mountain landscape, is referred to neither as a perceptual object in an act of PER x, nor as a purely represented object in one of the simpler forms of REP x, if instead the mountain landscape is referred to *in so far as it appears in a picture* y, then a novel situation obtains in which x, the mountain landscape, is *not actually perceived*, even though it as it were appears over there in the picture y that is actually perceived. The mountain landscape that is shown "in" the picture is *not believed to be really there* in front of the perceiver; in its way of appearing it is taken to be something *unreal*, an *image* only of something else that in itself may be real or purely fictional. Put another way, if the object x, the mountain landscape, is *not referred to as it is in itself*, but rather only in so far as it is depicted in the picture y, we have a peculiar novel entity, the pictorial object or the image (Husserl's *Bildobjekt*) which, as such, is taken to be *unreal*, or to be a mere semblance (*blosser Schein*); it will be designated as

$$\frac{-x}{y}$$

i.e. 'unreal x founded in, or carried by, y'. The pictorial object, the image of the mountain landscape, is the represented (depicted, displayed, portrayed, etc.) object x only in so far as it appears "in" the thing-like picture y. In image consciousness, the represented object x appears only *indirectly*, i.e. "in" something else, the picture y, that *directly* appears in correlation with someone's actually occupied perceptual point of view. This state of affairs would be unintelligible if the two objects

---

[18] See, e.g., Gregory (1970), chapter 2. "The peculiarity of pictures", p. 32; or Gregory (1974), section 54, "The speaking eye", p. 614ff.

(the pictorial object and the thing-like picture) had the same actuality-value (*Wirklichkeitswert*). But this, however, is precisely not the case. Intentional reference, e.g., to the depicted mountain landscape only occurs in so far as the unreal pictorial object, the image of the mountain, is perceived in a mode of non-actuality while the thing-like picture is taken for real.

When describing the pictorial object, mention of 'the object x' has been made, on the one hand, with regard to the *depicted* object and, on the other, the object x has been qualified by the adjective 'pictorial'. Such a way of speaking about the reflective finding may serve as a first hint at what has to be described as intentional reference to the *double object* 'pictorial object depicted object', as achieved by the twofold intentionality.

Using the notation, the structure of an act of pictorially representing x may be designated thus:

$$(2a) \qquad \underset{\text{(PRE)s}}{\text{i} \underline{\qquad}} \text{(REP} - \text{[PER])} - \underset{\text{(PER)}}{\underline{x}} \underset{y}{|- /-} \text{x}$$

This means: I, while grounded in actually presenting my surroundings s, am actually representing x, believed to be real or to be fictional, by means of representing (reproducing) a non-actual perceiving of x, in so far as it, taken to be unreal, appears in the picture y that I actually perceive.

The formula tries to mirror the reflective finding that the objective identity of the represented x itself is split into the double object 'pictorial object depicted object'. Each of the distinguishable parts on the objective side – 'x', '−x/y', 'y' – can be seen to be correlated with constituent parts within the unity of performing the representational act of PIC x:

$$(2b) \qquad \ldots \text{(REP } - \underline{\left[\text{PER}\right]}) \ldots \\ \ldots \text{(PER)} \ldots$$

If the question of reference is thus considered from the point of view of the intentionality of consciousness, *no part of the double object will occur without the others*.

The way of intentionally referring to an object x in PIC x can be summed up as follows. Without simultaneously referring to the x itself qua the depicted object x – whether x is a real or a fictional entity – reference to the pictorial object −x/y would not take place, except for aesthetical or theoretical reasons. Intentional reference may, however, still occur to the physical picture y alone in a simple perception or in an action of, say, taking the thing-like picture from the wall to put it somewhere else. In turn, without simultaneously referring to the pictorial object −x/y, there would of course be no reference to the x itself qua depicted object. However, in the absence of -x/y, reference to the x itself can occur, either in one of the simpler forms of purely mentally representing x, or in a direct perceptual referring to x. Neither of these forms of intentionally referring would entail a double object anymore. This much regarding image consciousness must suffice for present purposes.

Let us finally turn to somewhat more complicated forms of intentionally referring to an object in acts of intuitive representation. Up to now, the intentional property of "being implied in the mode of non-actuality" to be uncovered by phenomenological reflection has been confined to the activity of perceiving. However, the reflective finding of being implied in the mode of non-actuality extends beyond perception to mental acts in general, as soon as they are represented in a higher-order act. In "§112 of *Ideas I* (1913/1983), Husserl writes:

> He who is practiced in consciousness-reflections (and has previously learned to see any data of intentionality of whatever sort) will *see* precisely without any further difficulty (*eben ohne weiteres sehen*) the levels of consciousness which present themselves with phantasies in phantasies or with memories in memories or in phantasies. One will then see too what is inherent in the essential sort (*Wesensartung*) of this hierarchical formation: namely that *every phantasy of a higher level* can be freely converted into a *direct phantasy* of what was indirectly phantasied in it, whereas this free possibility does *not* take place in going over from *phantasy to the corresponding perception* (p. 263).

With more complex mental acts, such as those referred to in this passage, the formalism of the notation for expressing and surveying their inherent structures with regard to similarities and differences among them would seem to be particularly helpful. The more complex phenomena of intentional implication and modification belonging to distinctly different phenomenological forms of mental acts can, indeed, be succinctly captured by means of the formulae of the notation. In contrast, the sentences in the theoretical language of phenomenology that describe the phenomenological forms in question in, say, English will become more and more tortuous. Keeping in mind the preceding reflective descriptions of the structural components of the simpler forms of REM x and PIC x as expressed in the formulae (1b), (1c) and (2a), the following formulae may be presented without much further comment. Within the scope of this contribution, they only serve to give a glimpse of the potential of the formalism for more complex mental acts.[19]

To illustrate, consider the following cases of intuitive representational acts: (3) remembering to have remembered something – {REM REM x}; (4) imagining to be imagining a fictional object – {IMA IMA – x}; (5) remembering something that pictorially appeared in the past – {REM PIC x}; (6) merely imagining to be seeing something in a picture – {IMA PIC x}. By putting the formulae next to each other, similarities and differences come readily to the fore and can, using reflection upon the ways of givenness of the intentional objects, be understood as mirroring structural similarities and differences of the underlying representational acts.

(3) {REM REM x}, what is it "to remember having remembered something"?

(3a) $(\text{REP } p \vdash [\text{REP } p \vdash [\text{PER}]]) \vdash x$

(3b)   i _____ (REP p$|$– [_____ REP p'$|$–[PER]])$|$– x
         (PRE)s $|$–[PRE]s'

This means: I while grounded in the bodily presentation of my actual surroundings s am representing x, taken to be real, by means of representing (reproducing) another act of representing x, believed to have actually occurred in the past, such that, while quasi-grounded in the past surroundings s', believed to have actually been presented, a perceiving of x is represented (reproduced) to have been represented (reproduced) and is believed to have actually occurred in a more remote past'.

(4)   {IMA IMA – x}, what is it "to imagine to be imagining a fictional object"?

(4a)   i _____ (REP p$|$–[_____ REP – [PER]]) – x
         (PRE)s   – [PRE] $|$– / – s'

This means: I while grounded in the bodily presentation of my actual surroundings s am representing a fictional x by means of representing a neutralized act of representing x such that, while quasi-grounded in the represented surroundings s', either taken possibly to be real or itself fictional, a perceiving of x is represented to be neutrally represented.

   Husserl (2005) gives a nice example for an imagination of this sort: "I live in the imagining of a centaur. I imagine that I am carrying out this imagining. For example, I phantasy myself into the following situation: I am travelling in Africa; I rest from my march and give myself up to my phantasies; I think of centaurs and water nymphs in the world of the Greek gods." And he comments as follows: "These phantasies are not taken as present phantasies but as phantasies that are themselves phantasied. Within the phantasy, a distinction is again made between reality and dreams (phantasy)" (p. 220).

(5) {REM PIC x}, what is it "to remember something that pictorially appeared in the past?"
(6) {IMA PIC x}, what is it "merely to imagine seeing something in a picture"?

   Taking for granted what has been developed above for formula (2), formulae (5) and (6) may be written out as follows. In view of an example in PART II, only (6a) will be spelled out in English, describing for a change the formula on the basis of asking how the object intentionally referred to is given in IMA PIC x, rather than by asking what I am doing when representing something pictorially appearing in an imagination. It should be obvious how (5a), mutatis mutandis, would have to be spelled out in turn.

(5a)   i _____ (REP p$|$–[_____ REP – [PER])$|$– <u>x</u>$|$$|$– / – x
         (PRE)s $|$–[PRE]s' $|$–[PER]   y

(6a)
$$i \underline{\qquad} (REP - [\underline{\qquad} REP - \underline{[PER]}) |- \underline{x}||-/-x$$
$$(PRE)s \qquad -[PRE]s' \qquad -[PER] \quad y$$

The latter, (6a), means: some real or fictional object x is given to me in my actually representing x by means of a neutrally represented representing of x, such that, while quasi-grounded in a neutrally represented presentation of my surroundings s', a neutralized perceiving of x is represented in so far as x, taken to be unreal, appears in the picture y that is given to me by means of a neutrally represented perceiving of y while at the same time I am actually presenting my surroundings s.

Concerning all the formulae of the phenomenological notation, it is essential to keep in mind that for achieving the main goal of articulating varieties of intentionality with precise expressions, actual reflection upon the mental acts themselves and the ways of givenness of the objects intentionally referred to must always be kept in force. As Husserl observed in a research manuscript from around 1912 in which he explores various aspects of a phenomenological analysis of intuitive representations: "The phenomena of the *overlapping* and *mixing* of *intuitions* (*Deckung und Durchsetzung von Anschauungen*), and the experiences of agreement, conflict, and the modes of position taking belonging to them, must be studied in detail, since otherwise one easily goes astray".[20]

An Application to Scientific Studies of Consciousness

It seems clear and, probably, even uncontroversial that when scientists study the workings of the brain with the aim of looking for a scientific, ultimately a physical, explanation of consciousness, a clear conception of what they are seeking to explain is requisite. Here, Shaun Gallagher's (2003) proposal of a "front-loaded phenomenology", i.e. the idea of making direct use of phenomenology in the very design of empirical investigations of consciousness, presents itself. Along such a methodological line of integrating first- and third-person data, the argument is that, if you want a science of consciousness, look to the finer details of the conscious experiences themselves with the help of reflective phenomenology and aim at integrating this reflection-based knowledge with neuroscientific knowledge about the details of the brain. Within the scope of this contribution, referring to work on *mental imagery* and *the neural foundations of imagery* must suffice as an example to make plausible the view that the ever more advancing studies into the brain in the quest for consciousness should take advantage of what phenomenological clarifications of the very subject-matter can provide.

Kosslyn, in particular, and his collaborators (e.g., Kosslyn et al. 2001), thought that until recently issues concerning mental imagery had fallen within the purview

---

[20] See Edmund Husserl (2005), Appendix L, p. 569.

of philosophy and cognitive psychology where no substantial progress had been made, whereas with the advent of cognitive neuroscience questions about imagery have become empirically tractable, allowing theories of imagery to be tested objectively in humans (2001: 635). Taking advantage of developments in neuroimaging technologies, researchers have shown "that mental imagery draws on much the same neural machinery as perception in the same modality" (635). Indeed, approximately two-thirds of all the brain areas activated during perception and during imagery were found to be activated in both cases (636). These and other findings indicate that imagery and perception share very specific, specialized mechanisms. But the two, imagery and perception, do not draw on identical processes. As Kosslyn et al. put it:

> Although shape, location and surface characteristics are represented and interpreted *in similar ways* during both functions, the two *differ* in key ways: imagery, unlike perception, does not require low-level organizational processing, whereas perception, unlike imagery, does not require us to activate information in memory when the stimulus is not present (2001: 636; emphasis EM).

Imagery researchers, such as Kosslyn et al., conclude that "images are in fact *internal representations*" (641) "that *depict* information, not describe it" (639), and this they take to be "evidence that mental imagery relies on actual images" (639), evidence that seems mainly to be gathered from the activation of the early visual cortex and to be quite solidly supported by numerous imaging studies (640). However, from the point of view of a phenomenological analysis of the phenomena in question, the lack of appreciation, in this work, of fundamentally differently structured ways of intentionally referring to something in using imagery is quite apparent. Thus, Kosslyn et al. report that, e.g., *visualizing* an object has much the same effects on the body as *actually seeing* the object, or when subjects *view pictures* of the objects under study, e.g., threatening objects. Again, they say, *much the same effects* on the body occur, as recordings from single cells in the human brain have shown, "while subjects were shown pictures or formed mental images of those same pictures" (641). *Taken in isolation*, the findings of overlapping specific cortical areas in perception and imagery no doubt corroborate the view of an inner connection between perception and forms of imagination and picturing, showing "much the same effects on the body". However, our conscious experiences do occur as *unified experiences* containing a differentiated manifold of moments or components within themselves, only some parts of which seem to overlap whereas the concrete experiences, as the wholes they are, are lived through with a distinctly different consciousness regarding the objects given in their presence or absence. And relative to this side of the topic, namely to the modes of *consciousness* or *phenomenal intentionality* pertaining to the experiences that are involved in imagery, some more developed reflective phenomenology would seem to be called for.

In order further to clarify what some of the conscious experiences that appear to be involved in the neuroscientific investigations of mental imagery *possibly* consist in and how they may be lawfully related to one another, Husserlian phenomenology

along the lines presented in Part I should be helpful.[21] Within philosophical Phenomenology, the notation as presently developed offers formulae for *idealized* forms of consciousness, mirroring the conditions of the *possibility* of the mental acts in question. Thus, in particular, the formulae always simply contain the expression 'PER' without specifying which perceptual modalities be involved in *actual* occurrences of this or that experience. Moreover, nothing so far is provided in order to signal differences regarding one's more or less attentively referring to some represented object while simultaneously being still aware in some degree of one's present surroundings, etc. In view of an application of the notation in the cognitive neurosciences, it would however be a relatively straightforward task when designing experiments to supplement the expressions according to the given cases. In so making explicit, which perceptual modality is actually involved in, say, a given imagery task – e.g., visual or tactual perceiving or a multimodal combination – and/or how one's attention is distributed, more precision would be attainable in the search for the relevant neural correlates.

Special attention is again to be given to the phenomena of intentional implication or modification of experiences within the unified representational experiences of remembering, imagining, picturing, and combinations thereof. For the neuroscientific research in question, one of the most vital aspects of the search for neural correlates is that such correlates should answer to the question of the *binding problem*. Phenomenological clarifications of conscious experiences, in so far as they aim at making explicit lawful internal connections among the components making up those unified experiences, should prove to be particularly apt to shed light on the issue of the binding by synchronized neural firing and accordingly to play a heuristic role in designing neuroscientific research concerned with neural correlates of consciousness.

With the help of the notation, it can readily be shown that phenomenological first-person data can, on the one hand, be seen to corroborate the scientifically well-established view of an inner connection between perception and forms of imagination and picturing. On the other hand, however – the neuroscientific finding of "much the *same* effects on the body" *notwithstanding*, as measured in perception and when using imagery and/or pictures (see above) – phenomenological analysis of the relevant representational experiences crucially makes explicit that distinctly different ways of intentionally referring to some represented object in using imagery and/or pictures are involved.

---

[21] With reference to Marbach (1993), Varela (1996) and Van Gelder (1996–1997), Gallagher (1997) observes that they "abandon polemics and set out to show, in some precise details, how phenomenology, specifically the kind of phenomenology developed by Husserl, can contribute to and constrain cognitive scientific approaches to consciousness" (p. 197). Upon presenting bits of the notation introduced in Marbach (1993), Gallagher mentions Antonio Damasio's *Descartes' Error* (Damasio 1994) as "an interesting example of how phenomenology and cognitive neuroscience might enlighten each other", pointing out that "Marbach provides the phenomenological analysis that fits Damasio's observations about how the brain works in cases of memory; Damasio provides the neurological evidence that supports Marbach's claim" (p. 200, note 5).

Consider in turn along the lines of the work on mental imagery reviewed by Kosslyn et al.

(7) A case of simply actually seeing – 'PER' – an object x presently out there
(8) Visualizing – 'IMA' – the object using imagery
(9) Viewing a picture – 'PIC' – of the same object x
(10) Imagining (or remembering) the picture – 'IMA PIC', or 'REM PIC' – showing the same object x

Recall that when phenomenologically examining these cases, it will be helpful to ask oneself, how is the object x given in each one of these experiences of intentionally referring to x? The following four structurally clearly distinct formulae for the phenomenological forms of these experiences will then obtain:

(7) Actually simply seeing object x, reflectively yields:

$$(PER)\ x$$

i.e. x is given by means of actually perceiving (seeing, touching, etc.) x.

(8) Visually imagining, i.e. visualizing, object x: (IMA) x, reflectively analyzed, yields

$$i\ \underline{\hspace{1cm}}\ (REP\ p-[PER]\ )-/|-x$$
$$(PRE)\ s$$

i.e. some fictional or real object x is given to me in my actually representing x by means of representing (reproducing) a neutralized perceiving of x while at the same time my surroundings s are actually presented.

(9) Viewing a picture of object x, or having an image consciousness of x: (PIC) x, reflectively analyzed, yields

$$i\ \underline{\hspace{1cm}}\ (REP-[REP]-\underline{x}|_{-/-x}$$
$$(PRE)s\qquad (PER)\quad y$$

i.e. some real or fictional object x is given to me, or appears to me, in my actually representing x by means of representing (reproducing) a neutralized perceiving of x in so far as x, taken for unreal, appears in the picture y that I actually perceive while at the same time my surroundings s are actually presented.

Regarding the more complex formula for the experience of imagining a picture showing x, (10) IMA PIC x (see (6) in Part I), the formula may be developed in three steps in view of making the reflective analysis more transparent:

(10) Imagining to be picturing object x; or taking up Kosslyn et al.'s terms: forming a mental image of a picture of object x: (IMA PIC) x, reflectively analyzed yields

(10a) $$i\ \underline{\hspace{1cm}}\ (REP-[REP...])|_{-/-x}$$
$$(PRE)s$$

i.e. some real or fictional object x is given to me in my actually representing x by means of a neutrally represented (reproduced) representing of x while at the same time my surroundings s are actually presented.

$$(10b) \qquad i \underline{\quad\quad\quad} (REP-[REP-[\underline{PER}]])|_{-}\underline{x}||_{-/-x}$$
$$(PRE)s \qquad\qquad -[PRE] \qquad y$$

i.e. some real or fictional object x is given to me in my actually representing x by means of a neutrally represented (reproduced) representing of x, such that a neutralized perceiving of x is represented in so far as x, taken to be unreal, appears in the picture y that is given to me by means of a neutrally represented (reproduced) perceiving of y while at the same time my surroundings s are actually presented.

$$(10c) \qquad i \underline{\quad\quad\quad} (REP-[\underline{\quad\quad\quad} REP-[\underline{PER}]])-\underline{x}||_{-/-x}$$
$$(PRE)s \qquad -[PRE]s' \qquad -[PER] \qquad y$$

i.e. some real or fictional object x is given to me in my actually representing x by means of a neutrally represented (reproduced) representing of x, such that, while quasi-grounded in a neutrally represented (reproduced) presentation of my surroundings s', a neutralized perceiving of x is represented (reproduced) in so far as x, taken to be unreal, appears in the picture y that is given to me by means of a neutrally represented (reproduced) perceiving of y while at the same time my surroundings s are actually presented.

Now, as hinted above, it should, on the one hand, be visible from formulae (7)–(9) and (10c) that the expression 'PER', designating an activity of perceiving, recurs each time. In (7) as designating an actually occurring act of perceiving the object x, indicated by the parentheses '(…)'; in (8) and (9) and (10c), the expression 'PER' appears within a pair of brackets, '[…]', contained on the upper line of the formulae, indicating thereby that the activity of perceiving is only represented (or reproduced) as perceiving in the mode of non-actuality and *no longer as experienced as actually occurring*. In other terms, perceiving is involved as being intentionally implied or modified within the consciously experienced unity of actually establishing intentional reference to the represented object x. In some way, then, it is the "same" perceptual activity with the "same" objective phenomenal content – say, an object x in its surroundings appearing in such and such shapes and colors – that occurs in one's actually seeing x as well as in one's visually representing x in one way or another. However, it is clear that *experientially*, i.e. from the first-person perspective of my conscious experiences, the intentional reference to the object x is altogether differently characterized when I am actually seeing x as against only representing (reproducing) a seeing of x in one way or another. And these are much sharper differences between *perception and imagery* than the ones mentioned by Kosslyn et al. themselves concerning aspects of organizational processing and of activation of information (see above). The differences are epistemic, instead; they are crucial with respect to how someone takes a given object to be (e.g., real, fictional, in the past, in a merely imagined world, etc.), and how someone takes a

given represented (reproduced) perceptual (or any other) activity to be (e.g., believed to have occurred in the past, or merely imagined without belief nor disbelief, but neutrally, etc.). Such differences truly make a difference in our daily life.

The various formulae permit us to show succinctly that, besides the component 'PER', much else is also involved, making an experience of simply seeing an object distinctly different from any experience of representationally referring to the same object and making one kind of representational experience distinctly different from another kind. To round off these reflection-based remarks, consider, in particular, the component of the unified experience that is expressed as

$$\overline{(PRE)}s$$

and that appears in (8)–(10), all three being forms of representational experiences that arise, so to speak, out of a grounding activity of presenting one's actual surroundings s: (PRE)s. This contrast between some presentational and simultaneously occurring representational activity within a unified experience of representational consciousness of something is a crucial first-person phenomenological datum. The contrast can be found to be reiterated in (10), making that experience of visually referring to the object x all the more involved in contrast to the other experiences the formulae of which are shown in (8) and (9), respectively. Last but not least, a further crucial first-person phenomenological datum concerns the difference between (8) – Kosslyn et al.'s visualizing – and (9) as well as (10) – involving pictures – that can be gathered from the formulae by noting that in (9) and (10) the object referred to is not the object x as it were *itself*, as it is the case in (8), but rather the complex "double object" of '–x/y | |– / – x', i.e. the object x (the depicted object) in so far as it appears in the picture y (see above, Part I).

The suggestion now is that these and other reflective findings concerning first-person data should be systematically integrated with third-person neuroscientific data concerning conscious experiences that make use of imagery in one way or another. The first-person phenomenological data, understood as data that provide insight into the very conditions of the possibility of conscious experiences of this or that form, provide the basis for *phenomenological concepts* that help *describe* the phenomena to be *explained* scientifically with the help of neuroscientific experimentation. More specifically, the phenomenologically conceptualized data should be of heuristic use for determining more precisely which synchronous neural firings are involved in using imagery by means of representing (reproducing) in one way or another a perceiving of an object – for the *binding problem* is obviously not limited to the sphere of perceptual-phenomenal consciousness in actual perception.

Elaborating a little more in view of integrating first- and third-person-data, consider the following phenomenological constraints to be taken account of in the experimental work: In a case of simply visualizing an object x – see formula (8) – some neural activity should be identifiable using third-person data in correspondence to a person's consciously modified experience of *seeing as it were* that is implied in referring to the object x, such that the pattern would not only show an

objective overlap with a pattern of some actual seeing of x – see formula (7) – e.g., concerning data corresponding to the shape and color of object x. Instead, the pattern of neural firing would have clearly to differ in its overall shape, given the phenomenologically crucial difference between a person's *actually* experiencing something with reference to object x as against experiencing *only as it were* something with reference to the same object x in one's visually imagining the object while at the same time actually experiencing something with reference to one's present surroundings. Similarly again, but with some additional complications, if we were to contrast the case of visualizing the object x – see formula (8) – and the case of forming a mental image of a picture of object x – see formula (10c). The corresponding patterns of neural firings would have to differ radically from one another, given the radically different subjective experiences described above with the help of phenomenological concepts and mirrored in the formalism of the notation.

To conclude, gathering reliable first-person data concerning *conscious experiences as such* is clearly not a straightforward affair, neither within philosophical Phenomenology nor in applications to the science of consciousness. Ideally, in the latter respect, methodologically sophisticated phenomenologists should be included in research teams, providing in advance of an experiment the relevant reflectively gathered structural first-person data concerning possible conscious experiences. Researchers could then take the phenomenologists' descriptive reports of this or that kind of conscious experience as a heuristic guide for designing experiments that uncover the finer details of an experience, of which, based on a suitable instruction to a participant, one would have good reasons to believe that it could be an instance of this or that kind of experience actually lived through by the participant in the experimental situation.

The phenomenological descriptions provide conceptually-based constraints for the empirical work by making *explicit in advance* what is only implicit in the pre-reflective natural consciousness of participants. Probably the main advantage of such an approach is that detailed research questions concerning distinctly different ways of being conscious can be addressed using all available third-person methods *without interfering* either with participants' performance or the very collection of third-person data. Note the crucial difference between, on the one hand, a participant's actively being involved in a conscious experience of one kind or another – say, an experience of viewing a picture of object x – formula (9) – or an experience of forming a mental image of a picture of x – formula (10c) – and, on the other hand, a phenomenologist's reflectively describing the very structure of such an experience with the tools of the phenomenological methodology. Whereas the phenomenologist's job can be done prior to, and in fact quite independently of, a given experimental situation, the participant's conscious experiences will actually be lived through, but not reflected upon, while *simultaneously* being monitored and measured with the help of third-person methods of brain imaging via fMRI and PET technology, surface recordings through EEG and MEG, etc.

A further advantage of the present approach, besides avoiding impeding interferences with participants' performance and data collection, consists in making

replications of the experimental situation readily available. Moreover, perhaps in connection with replications, new questions regarding further details and refinements of a conscious experience may come to the fore, regarding, for example, participants' shifting attention, following a suitable instruction, from intentionally referring to object x to referring to its way of appearing in a picture, or participants' consciously modifying a belief-attitude to an attitude of merely imagining something, etc. Furthermore, the first-person structural knowledge concerning consciousness that Husserlian phenomenology is able to provide should also lead to more refined third-person data connected to layers involved in participants' representational experiences (first-order, second-order, etc.) and to their lawful dependencies.

Last but not least, the phenomenological clarifications and conceptual constraints, so different from more or less personal reports about one's experiences, should of course be combined with other available methods and measures, thus no doubt *also* with participants' retrospective reports about what they had been doing during the experimental situation. At any rate, trying to elaborate an adequate neuroscience of *consciousness* should no longer be undertaken while ignoring the potential of Husserlian phenomenology. With its help, there is a fair chance that the "major pro-gramme for twenty-first century science", that Chris Frith (2002) evoked, "to dis-cover how an experience can be translated into a report, thus enabling our experiences to be shared" will get closer to its realization. Perhaps, a more advanced notation with a properly developed syntax that is still lacking in the present contribution will one day truly advance our knowledge of the nature of consciousness.[22,23]

References

Borsato A (2009) Innere Wahrnehmung und innere Vergegenwärtigung. Epistemata Philosophie, Königshausen & Neumann, Würzburg

Chalmers D (1999) First-person methods in the science of consciousness. *Consciousness Bulletin*, Consciousness studies, Department of psychology, The University of Arizona, Tucson, pp 8–11

Damasio AR (1994) Descartes' error. Emotion, reason, and the human brain. G.P. Putnam's Sons, New York

Dennett DC (1991) Consciousness explained. Little, Brown, Boston, MA

Frege G (1882a) Über den Zweck der Begriffsschrift: Suppl. zur *Jenaische Zeitschrift für Naturwissenschaft*, 16 (Neue Folge 9), 1882/1883 [1–10]. Reprinted in Frege, Angelelli (ed) 1964, pp 97–106

---

[22] Without being able to do justice to them in the present version of the notation, a certain "affinity" in spirit of wanting to present a logically rigorous Husserlian Phenomenology as potentially valu-able in the context of contemporary discussion of consciousness is to be found in Dan Lloyd (2004) and Wayne M. Martin (2005).

Frege G (1882b) Über die wissenschaftliche Berechtigung einer Begriffsschrift. In: Angelellli (ed) Zeitschrift für Philosophie und philosophische Kritik. 81, pp 48–56. Reprinted in Frege, 1964, pp 106–114

Frege G (1964). Begriffsschrift und andere Aufsätze. Zweite Auflage. Mit E. Husserls und H. Scholz' Anmerkungen herausgegeben von Ignacio Angelelli. Wissenschaftliche Buchgesellschaft, Darmstadt

Frith C (2002). How can we share experiences? TRENDS Cogn Sci 6(9):374

Gallagher S (1997) Mutual enlightenment: recent phenomenology in cognitive science. J Conscious Stud 4(3):195–214

Gallagher S (2003) Phenomenology and experimental design. Towards a phenomenologically enlightened experimental science. J Conscious Stud 10:9–10

Gregory RL (1970) The intelligent eye. Weidenfeld & Nicolson, London

Gregory RL (1974) Concepts and mechanisms of perception. Charles Scribner's Sons, New York

Husserl E (1977) Phenomenological psychology. Lectures, summer semester 1925 (trans: John S). Martinus Nijhoff, The Hague

Husserl E (1983). Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. First book: General introduction to a pure phenomenology (trans: Kersten F). Martinus Nijhoff (originally published in 1913), The Hague

Husserl E (1984). Logische Untersuchungen. Zweiter Band, Erster Teil: Untersuchungen zur Phänomenologie und Theorie der Erkenntnis, Ursula P (ed). Martinus Nijhoff (originally published 1901), The Hague

Husserl E (1989) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. Second book: Studies in the phenomenology of constitution (trans: Rojcewicz R and Schuwer A). Kluwer, Dordrecht

Husserl E (1997) Thing and space. Lectures of 1907 (trans: Rojcewicz R). Kluwer, Dordrecht

Husserl E (2005) Phantasy, image consciousness, and memory (1898–1925) Edmund Husserl, Collected Works, Volume XI (trans: John BB). Springer, Dordrecht

Kern I (1975). Idee und Methode der Philosophie. Leitgedanken für eine Theorie der Vernunft. Walter de Gruyter, Berlin/New York

Kosslyn SM, Ganis G and Thompson WL (2001) Neural foundations of imagery. Nat Rev Neurosci 2:635–642

Kriegel U (2007). The phenomenologically manifest. Phenomenol Cogn Sci 6:115–136 (special issue on Dennett and Heterophenomenology)

Lloyd D (2004) Radiant cool. A novel theory of consciousness. A Bradford Book. MIT Press, Cambridge, MA

Marbach E (1993) Mental representation and consciousness. Towards a phenomenological theory of representation and reference. Kluwer, Dordrecht/Boston/London

Martin WM (2005). Husserl and the logic of consciousness. In: David WS, Amie LT (eds) Phenomenology and philosophy of mind. Oxford University Press, Oxford, pp 203–221

Metzinger T (1995) Editor's introduction. The problem of consciousness. In: Thomas M (ed) Conscious experience. Schöningh/Imprint Academic, Thorverton/Padeborn, pp 3–37

Noë A (2007) The critique of pure phenomenology. Phenomenol Cognit Sci 6:231–245 (special issue on Dennett and Heterophenomenology. Guest Editor: Alva Noë)

Roy J-M (2007) Heterophenomenology and phenomenological scepticism. Phenomenol Cogn Sci 6:1–20 (special issue on Dennett and Heterophenomenology)

Schmicking D (2005) Is there imaginary loudness? Reconsidering phenomenological method. Phenomenol Cogn Sci 4:169–182

van Gelder T (1996/1997) Wooden iron? Husserlian phenomenology meets cognitive science. In Jean P, Francisco JV, Bernard P, Roy J-M (eds) Naturalizing phenomenology. Issues in contemporary phenomenology and cognitive science. Stanford University Press, Stanford (1999), pp 245–265

Varela FJ (1996) Neurophenomenology: A methodological remedy for the hard problem. J Conscious Stud 3(4):330–349

Yoshimi J (2007) Mathematizing phenomenology. Phenomenol Cogn Sci 6:271–291

Zahavi D (2007). Killing the straw man: Dennett and phenomenology. Phenomenol Cogn Sci 6:21–43 (special issue on Dennett and Heterophenomenology)

Part II
Consciousness, Attention, and Emotion

Consciousness

Mark Rowlands

The Natural Attitude

The phenomenological study of consciousness is the study of consciousness from the *inside*. The simplicity of this definition is, however, merely apparent. The idea of studying consciousness from the inside is amenable to two distinct interpretations. One of these interpretations is almost universally presupposed in recent influential scientific and philosophical treatments of consciousness outside of the phenomenological tradition. The other interpretation forms the conceptual core of the phenomenological approach but has, for the most part, been curiously overlooked in those recent scientific and philosophical treatments.

When recent influential scientific and philosophical treatments of consciousness talk of studying consciousness from the inside, they do so through the prism of what Husserl called the *natural attitude*. From the phenomenological perspective, on the other hand, the idea of studying consciousness from the inside is constituted precisely by the rejection of the natural attitude. The two interpretations are, therefore, irreducibly distinct. They are not, necessarily, incompatible: arguably, both may be required for a complete understanding of consciousness. However, the distinctness of the interpretations does mean that recent scientific and philosophical treatments of consciousness have overlooked something crucial to consciousness. Indeed, I shall argue that they have overlooked what is most important about consciousness.

The natural attitude is a type of epistemic *stance* that we adopt. However, it is a stance that we can adopt not only to the external world, but also to our mental states:

M. Rowlands
Department of Philosophy, University of Miami, Coral Gables, FL, 33124

> Let us suppose that in a garden we regard with pleasure a blossoming apple tree, the freshly green grass of the lawn, etc. It is obvious that the perception and the accompanying liking are not, at the same time, what is perceived and liked. In the natural attitude, the apple tree is for us something existing in the transcendent realm of spatial actuality, and the perception, as well as the liking, is for us a psychical state belonging to real people. (Husserl 1983: §88)

This is an entirely typical Husserlian explication of what he means by the natural attitude. The attitude can be adopted not only toward the "apple tree" but also toward the "perception" of the apple tree and other types of "psychical state". Moreover:

> By the real phenomenological content of an act we mean the sum total of its concrete or abstract parts.… To point out and describe such parts is the task of pure descriptive psychological analysis operating from an empirical, natural-scientific point of view. Such analysis is in all cases concerned to dismember what we inwardly experience as it in itself is, and as it is really given in experience. (Husserl 1970: §16)

Notice here that Husserl claims that the "natural-scientific" is concerned with what we "inwardly experience … as it is really given in experience". In the next paragraph, Husserl explains in what rejecting the natural attitude and replacing it with a phenomenological one would consist:

> Let us now shift from our natural-scientific, psychological standpoint to an ideal-scientific one. We must exclude all empirical interpretations and existential affirmation, we must take what is inwardly experienced or otherwise inwardly intuited (e.g. in pure fancy) as pure experiences, as our exemplary basis for acts of Ideation.… We thus achieve insights in a pure phenomenology which is here oriented to *real* (*reellen*) constituents, whose descriptions are in every way "ideal" and free from "experience" i.e. from presuppositions of real *existence* (1970: §16)

This is an (early) description of the phenomenological reduction: the *epoché* or *bracketing*. When we perform the *epoché*: "The transcendent world receives its 'parenthesis,' we exercise the *epoché* in relation to 'positing' its actual being." (1983: §88). This sort of characterization of the *epoché* has given rise, in some circles, to the laughable idea that once you have suspended belief in the external world, and are focusing on things as they appear to you, you are doing phenomenology. But this is, at most, the starting point of phenomenology, not what phenomenology is.

Phenomenology is the attempt to study not what appears to consciousness, but to use what appears to consciousness as a starting point: to work back from what appears to consciousness and use this to understand those features of consciousness that allow what appears to consciousness to appear in the way that it does. In other words, the starting point for a phenomenological study of consciousness is performance of the *epoché* with regard to the items we encounter – whether these items are apple trees *or* experiences. Phenomenologically, to study consciousness from the inside is to study it not as an object of experience, even when this experience is directed towards conscious items and their properties. Rather, it is to study consciousness as that which permits objects of consciousness to appear (to subjects) in the way that they appear.

Recent scientific and philosophical treatments of consciousness are typically framed in terms of a distinction between *first-* and *third-person perspectives*.

The third-person perspective concerns, in the first instance, items possessed by another person. Scientific treatments of consciousness are based on the adoption of a third-person perspective. Here, conscious items are regarded as objects of scientific investigation. The precise nature of these objects will depend on the nature of the investigation in question. Neural processes are one common candidate, when the investigation is pitched at an appropriate level. Computational roles or algorithms are another candidate when the investigation switches so a slightly more abstract level of description. In this version of the attitude, conscious items are conceived of as objects of *thought* (broadly construed to include cognitive processes such as *theorizing*). This is one version of the natural attitude towards consciousness.

There is, however, another way of adopting the natural attitude towards conscious items. This occurs when one regards these items as objects of *inwardly directed awareness*, of *introspection*, where this is construed broadly enough to incorporate both inwardly directed experience and inwardly directed thought (and other propositional attitudes). From this *first-person* perspective, conscious items are the things one encounters when one turns one's awareness 'inwards': that is, when one's attention is appropriately engaged. As such, they can be either *experiential* objects (sensations, experiences, etc.) or *propositional* objects (thoughts, beliefs, desires, etc.) – depending on whether this awareness is experientially or propositionally based (this is not intended to preclude the claim that experience is itself propositional).

The important differences between first- and third-person perspectives should not blind us to what they have in common. The third-person, scientific, version of the natural attitude, at least in its paradigmatic form, attempts to make the mental objects of *others* into objects of investigation. The first-person, *introspective* version, attempts to make the mental objects possessed by oneself into objects of awareness. However, both are versions of the natural attitude because they are united by a common assumption: conscious items are *objects* of appropriately directed mental acts.

The essence of phenomenology lies in this idea: there is more to consciousness than can be made into an object of a mental act, whatever the specific character of this act. Therefore, *there is more to consciousness than what appears to consciousness*. To suppose otherwise is to think of consciousness only from the *outside*: as an extrinsic object of an intentional act. Understood from the outside, consciousness is one of the items revealed by the natural attitude – whether this takes scientific *or* introspective forms. Understood from the *inside*, consciousness is not what is *revealed* from the natural attitude: it is the *adopting* of (among other things) the natural attitude. That is, consciousness consists (also) in the adopting of the attitude from which it – among other things – can appear as a collection of conscious items (i.e. as objects of consciousness). That this is so, I shall argue, follows, ultimately, from the essential feature of consciousness: its *intentionality*. Before we turn to this, however, it is useful to consider the extent to which recent influential treatments of consciousness have overlooked this basic phenomenological insight.

The Pull of Objectivity

Almost all recent developments of the problem consciousness is thought to pose for materialist accounts of the mind – the *hard problem* as it is sometimes called – presuppose, sometimes implicitly but usually explicitly, that consciousness is an *object* of some sort. By this I do not mean, of course, an object as opposed to some other category of existent – event, state, process, property, fact, etc. Rather, consciousness is conceived as something *of* which we are aware in the having of experience. I shall refer to this as an *objectualist* account of consciousness (Rowlands 2001, 2002). According to objectualism, consciousness is an *object of awareness*. The specific nature of this object varies, but the following categories are important ones:

1. Consciousness is an object of *knowledge*.
2. Consciousness is an object of *introspection*.
3. Consciousness is something to which we have *access*.

These claims are not necessarily incompatible (although on certain interpretations they can be). But neither are they synonymous.

Frank Jackson's (1982, 1986) *knowledge argument* is explicitly predicated on claim (1). Colin McGinn's (1989, 1991, 2004) defence of his transcendental naturalist position explicitly requires claim (2). Thomas Nagel's (1974, 1986) position is, however, particularly instructive, because here we find an implicit commitment to (3), and this, in the work of the arch-champion of subjectivity, brings out just how widespread and tenacious is this *objectualist* conception of consciousness.

In his seminal (1974) paper 'What is it like to be a bat?' Nagel argued that (1) 'Fundamentally, an organism has conscious mental states if and only if there is something that it is like to be that organism – something that it is like *for* the organism.' (1974: 16). However, (2) 'If physicalism is to be defended, the phenomenological features of experience must themselves be given a physical account.' (1974: 167) But (3) 'When we examine their subjective character it seems that such a result is impossible. The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective physical theory will abandon that point of view.' (1974: 167)

What is important for our purposes is not the success or otherwise of Nagel's argument, but a particular conception of subjectivity embodied in it. Nagel begins with a certain common understanding of objectivity. An 'objective fact *par excellence*' is 'the kind that can be observed and understood from many points of view' (1974: 172). Objective facts are ones to which there exist many *routes of epistemic access*. It is the existence of such many and varied routes, capable of being adopted by many and varied individuals, that constitutes an item as objective. In short, objective items are ones to which epistemic access is *generalized*. Taking this concept of objectivity as primary, Nagel then constructs a concept of subjectivity based on the guiding metaphor of a route of access. Subjective phenomena are ones to which our routes of access are reduced to one: they are items to which our access is *idiosyncratic*. To think of subjective phenomena in this way is to think of them

as part of a region of reality that *in itself* is just like any other. This region of reality differs from other regions not in any of its intrinsic features: the only difference lies in our mode of access to it. Our port of epistemic entry to this region of reality is unusually small. Classically objective phenomena are like objects on a savannah, and can be approached from many different directions. Conscious phenomena are locked up in a remote canyon whose only route of access is a narrow tunnel.

This way of thinking about consciousness is, I think, part of the pull of the idea that all reality is intrinsically objective. Reality reduces, ultimately, to a collection of objects of consciousness. This is just one more expression of the natural attitude. As a result, objectivity is taken as primary, and subjectivity is understood as a derivative and truncated form of objectivity. And the notion of a mode or route of access lies at the heart of both concepts. Thus, it is our having idiosyncratic (i.e. truncated) access to an item that constitutes that item as subjective. If only our routes of access could somehow be *beefed up*; if only they could be suitably generalized, then the very same item would become objective. The idea that reality is intrinsically objective is the idea that this generalizing of routes of access could, in principle, take place without any change in the intrinsic nature of the object toward which this access is directed.

To see the significance of this way of understanding the subjective-objective distinction, consider Nagel's tendency to slide from claims such as:

> Every subjective phenomenon is essentially *connected* with a single point of view (1974: 167, emphasis mine).

To claims such as:

> For if the facts of experience – facts about what it is like *for* the experiencing organism – are *accessible* only from one point of view, then it is a mystery how the true character of experiences could be revealed in the physical operation of that organism (1974: 172).

The claim that a subjective phenomenon is one essentially *connected* with a single point of view mutates into the claim that a subjective phenomenon is one that is essentially *accessible* from only a single point of view. However, these two claims are not equivalent. To suppose that they are is a symptom of the grip exerted on us by the natural attitude that understands reality as being reducible to a set of objects of consciousness.

To think of subjective phenomena as items that are *accessible* is to think of them as objects of that access, and therefore as objects of consciousness. The idiosyncratic character of this access is one way in which their subjectivity might be explained. However, there is an alternative explanation of subjectivity; one that the ubiquity of the objectualist conception of subjectivity has rendered almost invisible, at least in recent treatments of consciousness. Subjective, conscious, phenomena are not parts of a region of reality to which our access is idiosyncratic, and where this idiosyncrasy constitutes their subjectivity. Indeed, they are not parts of a region of reality to which we have access at all. Rather, subjective, conscious, phenomena are ones that belong only to the *access* itself. There is no region of reality to which conscious phenomena belong, or in which they find their place. Rather, conscious phenomena simply belong to our *accessing* of regions of reality – regions that are,

in themselves, perfectly objective. It is this, crucial, phenomenological insight that has been lost in almost all recent treatments – both philosophical and scientific – of consciousness.

Consciousness as Empirical and as Transcendental

To think of consciousness as an object of awareness is to think of it as a collection of events or states, each possessing a constellation of properties. To think of consciousness in this way is to think of it as a collection of *empirical* things (Rowlands 2003). I use this term in a recognizably Kantian sense. To say that an item is empirical is simply to claim that it is an actual or potential object of consciousness: it is the sort of thing of which I might become aware if my awareness is suitably engaged. It is not only experiences but also their *properties* or *aspects* that can be empirical in this sense. I can turn my introspective attention towards, for example, my pain; but I can also turn my attention to its various qualities – whether it is throbbing rather than continuous, whether it is a dull and aching rather than sharp and stabbing, etc. In the empirical conception of consciousness, both my conscious states and their properties are objects of awareness – things towards which I can turn my attention.

The basic insight that underwrites the phenomenological study of consciousness is that, whatever else is true of it, consciousness is *necessarily* more than a collection of empirical things. To begin with, consciousness is essentially *hybrid*. Consciousness can be both *act* and *object* of experience. Consciousness is both the directing of awareness and it *can* be that upon which awareness is directed (Rowlands 2001, 2002). Consciousness is both the act of conscious experience, and it can be experience's object. However, steeped as we are in the natural attitude, a popular response to this (obvious) fact is that it is possible to make consciousness as act of awareness into an object of awareness. Whenever consciousness functions as an act of awareness it is always possible, in principle, to make this act into an object of awareness.

To take just one example of how this might go, consider a scientific version of the natural attitude that reduces visual perception to a series of computationally specified processes occurring in the brain. On David Marr's account, for example, we would functionally decompose visually perceiving the world into a series of discrete operations beginning with the retinal image, proceeding through the brain's successive construction of the raw primal sketch, the full primal sketch, the 2½D sketch and the 3D object representation. This is a way of turning the *act* of visually perceiving the world into an *object* of theoretically mediated awareness. The essential insight of phenomenological approaches to consciousness can, now, be put like this: necessarily, there is always as aspect of consciousness as act of awareness that resists the attempt to transform it into an object of awareness. In opposition to the empirical conception of consciousness, the phenomenological approach insists that there is an aspect of consciousness that is irreducibly *transcendental*: it is not something

of which we are aware, but something *in virtue of* which we are aware. Transcendentally, consciousness is something in virtue of which the objects of which we are aware appear to us in the way they do.

In the remainder of this paper, I shall defend this transcendental-phenomenological conception of consciousness. I shall argue that conscious experience contains what I shall call a *non-eliminable intentional core*. This core of any experience cannot be made into an object of awareness. And the existence of this core follows, ultimately and perhaps rather surprisingly, from the fundamental character of intentionality.

The Intentional Core of Experience

It is not only in the phenomenological tradition but also in the wider philosophical tradition that intentionality is understood as more than a simple directedness-towards an object. Rather, the concept of intentionality has usually been taken to embody a quite specific proposal about how this directedness is brought about. Intentional directedness is to be explained in terms of a tripartite structure of (1) act, (2) object, and (3) mode of presentation. The intentional act is connected to the intentional object via a mode of presentation of that object. Thus, a subject, in virtue of its intentional act, is aware of an object, and the act makes the subject aware of this object because it is this object that satisfies the mode of presentation embodied in the act. The mode of presentation is what allows the intentional act to 'hook onto' the intentional object of that act. I shall refer to this as the *mediational* conception of intentionality. If we adopt this mediational conception of the intentional relation, then the relation between an experience as act, E_A, and an experience as object, E_O, is this: E_A presents E_O to subject S by way of a mode of presentation, P, of E_O.

This is the way the concept of intentionality has been understood in the phenomenological tradition. In the wider philosophical tradition, the claim that all forms of intentional directedness must conform to this model has come under attack in recent decades, largely due to the groundbreaking work of Kripke (1980). This paper does not dispute the possibility of forms of intentionality that do not conform to the traditional model. However, it assumes that at least some forms of intentional directedness do thus conform. This is a common assumption, and the model of intentionality implied in it is sufficiently pervasive to be designated the *traditional* model of intentionality.

According to this traditional model, the mode of presentation is what connects intentional act to intentional object. Employing a terminology developed by Kaplan (1989), we can say that the intentional act has a *character* whose *content* can be expressed in the form of a description, and the intentional *object* of the act is the object that satisfies this description. The mode of presentation of the object, then, consists in the content expressed in the relevant description.

If an object satisfies the content-specifying description, however, this will be because the object possesses certain *aspects*: aspects that are picked out by the

content expressed in the description. Aspects are not to be identified with objective *properties* of objects. Aspects are objects of awareness in an intentional rather than objective sense. Aspects are the ways in which objects are presented, the ways in which they appear, to subjects. And to the aspect there may or may not correspond an objective property of the object. An object may appear – be presented as – round, even if it, in fact, is not. A necessary condition of an object having aspects (although not, of course, properties) is the intentional activity of a subject.

Since the *aspects* of the object are that in virtue of which it satisfies the content-specifying description, and since the mode of presentation of the object is the content expressed in that description, this invites the almost irresistible identification: we identify the *mode of presentation* of the object with that object's *aspects*. This identification, however, is problematic: it can be both true and false, depending on how we understand the notion of a mode of presentation.

Aspects are intentional objects of awareness. I can attend not only to the tomato, but to its size, colour and lustre. Indeed, it is arguable that I attend to the tomato in virtue of attending to these sorts of aspects. Thus, if we identify modes of presentation with aspects, and if we adhere to the traditional conception of intentionality as a relation whereby an object of awareness is determined only by a way of a mode of presentation, it follows that whenever there is a mode of presentation there must be another mode of presentation to fix reference to it. And if we make this second mode of presentation into an object of awareness – an *aspect* of our experience of which we are aware – there must be another mode of presentation that enables us to do this.

This is an issue of *non-eliminability* rather than *regress*. It is not that any experience must contain an infinite number of modes of presentation. That regress is stopped as soon as we stop trying to makes modes of presentation into objects of our awareness. For example, if we identify a mode of presentation of a tomato with an aspect of that tomato, and so think of it as an intentional object of my experience, then it follows from the traditional mode of intentionality that there must be another mode of presentation that allows it to be such. However, as long as I do not attempt to make this further mode of presentation into an intentional object, there is no need for an additional mode of presentation to fix reference to it. Therefore, in any given experience, there must be a mode of presentation that cannot, in that experience, be made into an intentional object. In the experience, the mode of presentation is not something *of* which we are aware (as we might, for example, be aware of aspects) but something *in virtue of which* we are aware of the intentional object of our experience.

The concept of a mode of presentation is, therefore, ambiguous. There is a consistent and respectable interpretation of that concept that thinks of it as, in essence, an aspect of objects. Thus, Husserl (1983) argues that any intentional act comprises a *noesis* and a *noema*. The Husserlian *noema* is typically understood as the way in which an object is presented to a subject. On this reading, the intentional *noema* corresponds to an aspect of an object. However, the traditional understanding of the intentional relation – endorsed by Husserl – entails that every intentional object requires a mode of presentation. Since Husserl claims that it is the experiential

noesis that allows the *noema* to appear as an object of an intentional act, this would suggest the identification of mode of presentation with experiential *noesis* rather than *noema*. The concept of a mode of presentation, therefore, seems to have a foot in both *noetic* and *noematic* camps. This ambiguity stems not from ambiguity in the Husserlian concepts of *noesis* and *noema*, but, rather, from the concept of a mode of presentation. This ambiguity, I have argued, can be traced to the traditional model of intentionality. This is not a failing of the traditional model, but it is nonetheless something that we must always bear in mind when we try to understand the nature of experience.

We can record this ambiguity in the concept of a mode of presentation by way of the neo-Kantian terminology introduced earlier. The concept of a mode of presentation admits of both *empirical* and *transcendental* interpretations. To say that an item is empirical is simply to claim that it is an intentional object, an actual or potential object of consciousness: it is the sort of thing of which I might become aware if my awareness is suitably engaged. Aspects of objects are empirical in this sense. A transcendental item, on the other hand, is one that is not and cannot be an intentional object – at least not in its transcendental role – because it is that which permits objects to appear under aspects. That is, in its transcendental role, a mode of presentation is a *condition of possibility* of intentional objects (although not, of course, of objects). An empirical mode of presentation is an aspect of objects. A transcendental mode of presentation is what makes a given empirical mode of presentation *possible*. This, ultimately, is what justifies the rubric *transcendental*. (Alternatively, if one does not like the appeal to empirical and transcendental interpretations of the concept of a mode of presentation, and would like to keep this concept univocal, it is possible to reject the idea that aspects are modes of presentation, and reserve the latter expression for that which fixes reference to aspects. For our purposes the consequences of this alternative approach are identical to the one developed here).

If we assume that the identification of modes of presentation with aspects is one legitimate way of understanding this concept, then the traditional model of intentionality has this clear entailment: any given experience must contain not only an empirical but also a transcendental mode of presentation. In this transcendental mode of presentation we find the *non-eliminable intentional core* of the experience. This is not an intentional object of the experience, and in its transcendental role cannot be an object of any experience. It is that which allows mundane worldly objects to be presented to subjects by way of aspects.

Intentionality, Body, and World

The idea of the non-eliminable intentional core of experience as transcendental mode of presentation allows us to properly understand the emphasis the phenomenological tradition has placed on what we might call the *situated* character of consciousness and intentionality: the role played by *bodily* and *worldly* structures in the constitution of conscious experience (Rowlands 1999).

A transcendental mode of presentation is what allows an object to be presented by way of an empirical mode of presentation. As such, we might think of transcendental modes of presentation as a type of *disclosing activity*. If I have a visual experience as of a shiny, red, tomato, then the transcendental mode of presentation of the experience is that which allows the tomato to be presented to me as red and shiny. This is true, of course, only if there is a tomato there. In the case of an illusory experience as of a shiny, red tomato, then it is still true that some part of the world – that which is erroneously taken to be a tomato – is presented as shiny and red. Again, it is the transcendental mode of presentation of the experience which allows the world to be presented in this way. The transcendental mode of presentation of my experience is that in virtue of which the tomato, or relevant part of the world, is *disclosed* or *revealed* to me as red and shiny. The non-eliminable core of intentional experience, therefore, consists in a *disclosure* or *revelation* of the world. The fundamental sense in which intentional acts are directed towards the objects, therefore, is that they reveal or disclose them as having certain *aspects* or *empirical modes of presentation*. This, I shall now argue, has one crucial implication: as directedness toward objects, intentional acts are also, necessarily, a *living-* or *passing-through* their material realizations.

This idea can perhaps best be clarified, in the first instance, by way of Merleau-Ponty's (1962) famous discussion of the perceptual role played by a blind person's cane (cf. Polanyi 1962). As Merleau-Ponty notes, it is possible to tell two quite different stories about this role. The first story treats the cane as an empirical object in the sense introduced above. The resulting empirical story is a familiar one. Tactile and kinaesthetic sensors in the blind person's hands send messages to the brain. Various events then occur in the person's sensory cortex, and these are interpreted as the result of ambient objects standing in certain relations to the person's location. When suitably filled out, there is nothing wrong with this story. However, it only describes the blind person's consciousness from the outside; as an empirical phenomenon. The story from the inside – the transcendental story in the sense introduced earlier – is quite different. The cane – in conjunction, of course, with the requisite neural and other biological machinery – *discloses* or *reveals* objects as possessing or falling under certain *aspects* or *empirical modes of presentation*. Thus, an object may be disclosed to the blind person as being 'in front' of him or her, as 'near', 'further away', 'to the left', 'to the right', and so on.

Merleau-Ponty is at pains to emphasize – correctly – the phenomenology of the resulting perception of the world. The blind person does not experience aspects of the objects he encounters as occurring in the cane, even though this is (part of) the material basis of his perception of these aspects. Still less does he experience them as occurring in the fingers that grip the cane; and less again in the sensory cortex that systematizes the experiential input. There is an empirical story to tell of why this is so; essentially a story of inference. The brain infers the location of aspects from the information it receives through its sensory apparatus. There is nothing wrong with this story. However, there is also another story to be told. Transcendentally, aspects are experienced as features of the world, and not as modifications of the material realizations of our consciousness, because consciousness passes through

these material realizations all the way out to the world. From the inside, experience does not stop short of the world, and that is why it can be experience *of* the world.

If the central contention of this paper is correct, then this phenomenological feature of experience is grounded in the fundamental nature of intentionality as disclosure. Intentional acts are directed towards the world in the sense that they are revealing activity. But where does the blind person's revealing activity occur? When the person discloses an object as being in front of him, for example, where does this disclosing activity occur? It occurs, in part, in the brain. But it also occurs in the body, and also, crucially, in the cane and the cane's interaction with the world. Revealing activity, by its nature, *does not stop short of the world:* it *passes through* its material realizations out to the world itself.

The role of the cane, fundamentally, is not one of *object* of disclosure but, rather, of *vehicle* of disclosure. The blind person does not experience the object as 'on the end of the cane', nor does he experience it as a blocking or resistance to the cane. Rather it is *in virtue of* the object being on the end of the cane, and *in virtue of* the resistance it provides to the cane, that the blind person experiences the object as spatially located in the world. In employing the cane, the blind person ceases to experience the cane. The cane becomes a vehicle of his experience not an object of it. As revealing activity, his experience passes all the way through the cane to the object itself. That is why his experience is *capable* of being a disclosing of the aspects of those objects.

The sort of disclosure afforded by the blind person's cane is by no means idiosyncratic. Consider the disclosing activities of a visually unimpaired subject. It is tempting to think of these activities as restricted to processes occurring in the eyes themselves, and subsequent neural processing operations. Such processes would, of course, be vehicles of disclosure, not objects of disclosure. And if the perceptual revealing activities of an unimpaired subject were restricted to these, then one would be constrained to think of these activities as stopping short of the world. However, the subject's disclosing activities consist in far more than these processes. For example, saccadic eye movements are an important component of the visual disclosure of the world. When I perform visual tasks, my eyes engage in various movements or *saccades*. Yarbus (1967) has demonstrated that the pattern of saccadic eye movement is systematically related to the nature of the task. Yarbus asked subjects, prior to their viewing of a painting, to perform certain tasks. The painting showed six women and the arrival of a visitor. Subjects were asked to either:

1. View the picture at will.
2. Judge the age of the people in the painting.
3. Guess what the people had been doing prior to the arrival of the visitor.
4. Remember the clothing worn.
5. Remember the position of objects in the room.
6. Estimate how long it had been since the visitor was seen by the people in the painting.

Yarbus demonstrated that different tasks resulted in quite different scan paths. Subjects asked questions concerning the appearance of people in the painting – for example, questions about their ages – focused on the area around the face. Subjects

asked questions concerning the theme of the painting focused in points throughout the picture. And different themes resulted in different scan paths. For example, subjects asked what the people in the painting were doing before the visitor arrived employed a different scan path from those asked to estimate how long it had been since the visitor was last seen by the family. In general, Yarbus showed, the scan varies systematically with the nature of the task.

From the outside, as empirical processes, my saccadic eye movements may be seen as part of a hypothesis testing apparatus. For example, the brain forms a hypothesis concerning the dispositions or activities of the people in the painting, and the saccades are, in part, what allows this hypothesis to be tested. As such, saccades may be understood – I think questionably – as external aids in the formation of conscious experiences. However, from the inside, as transcendental revealing activity, my consciousness lives or passes through my saccadic movements. The scan path is part of the material realization of my disclosing activities: it is a vehicle of disclosure – part of what allows me to disclose the world as being one way rather than another (e.g. as a world where a group of people were doing X prior to the visitor's arrival rather than Y). As such, it is one of the means by which my consciousness passes all the way out to the world. As directedness towards the world, my visual consciousness is also a passing through the saccadic scan path.

As examples of the same general sort, consider the probing, exploratory, activities I perform on the world when I identify its sensorimotor contingencies. If I direct my attention at will to various aspects of the visually presented world, or if low-level attention-grabbing mechanisms automatically direct my attention towards a visual transient, these activities are all part of the material basis of my visual perception of the world (and the saccadic scan path is, in effect, one component of these activities). Without these activities, my visual perception of the world would be very different and, at best, significantly attenuated. These activities allow me to disclose the visual world as, for example, complex, detailed, and stable (O'Regan and Noë 2001; Noë 2004). However, when I experience the world as being this way, I do not experience the activities in virtue of which it appears to me as such: these activities are vehicles of my disclosure not objects of it. From the inside, as transcendental mode of presentation, my consciousness passes through these activities and does not stop short of the world. From the inside, my consciousness lives through my probing, exploratory, activities in much the same way that the consciousness of the blind person lives through his cane.

In this respect, a useful template for thinking about revealing activity is provided by processes such as *exploration*. Exploration is a process that occurs in the head, in the body, and also in the world. By its very nature, exploration of a given region of the world does not stop short of that world: if it did it would by definition, be unsuccessful (or even fail to be a process of exploration at all). Exploration is, in this sense, essentially *worldly*. Revealing activity is worldly in roughly the sense that exploration is worldly, and is so for the simple reason that exploration is one form that revealing activity can take (Rowlands 2006).

Conclusion

This paper has defended the following claims: (1) phenomenology is the attempt to study consciousness from the *inside*, not as an object of awareness – whether third- *or* first-person awareness – but as that which permits objects to be revealed under aspects or empirical modes of presentation; (2) the non-eliminable intentional core of experience, therefore, consists in consciousness as transcendental mode of presentation; (3) as such, consciousness is essentially revealing or disclosing activity; (4) as revealing activity, consciousness is not only a directedness-towards objects; it is also a living or passing through of its material realizations to the world – indeed, it can only be the former because it is also the latter; and (5) as a passing through its material realizations, consciousness is indifferent to the location of those realizations.

References

Husserl E (1970) Logical investigations (trans: Findlay J). Routledge, London

Husserl E. (1983) Ideas pertaining to a pure phenomenology and phenomenological philosophy (trans: Kersten F). Martinus Nijhoff, The Hague

Jackson F (1982) Epiphenomenal qualia. Philos Q 32:127–132

Jackson F (1986) What Mary didn't know. J Philos 83:291–295

Kaplan D (1989) Demonstratives. In: Almog I, Perry J, Wettstein H (eds) Themes from Kaplan. Oxford University Press, Oxford

Kripke S (1980) Naming and necessity. Harvard University Press, Cambridge, MA

McGinn C (1989) Can we solve the mind–body problem? Mind 98:349–366

McGinn C (1991) The Problem of consciousness. Basil Blackwell, Oxford

McGinn C (2004) Consciousness and its objects. Oxford University Press, New York

Merleau-Ponty M (1962) The phenomenology of perception (trans: Smith C). Routledge, London

Nagel T (1974) 'What is it like to be a bat?' Philos Rev 83:435–450. (Reprinted in Nagel (1979) Mortal questions. Cambridge University Press, New York. All page references are to the latter).

Nagel T (1986) The view from nowhere. Oxford University Press, New York

Noë A (2004) Action in perception, MIT Press, Cambridge, MA

O'Regan K, Noë A (2001) A sensorimotor theory of vision and visual consciousness. Brain Behav Sci 24: 939–1031

Polanyi M (1962) Personal knowledge. Routledge, London

Rowlands M (1999) The body in mind: understanding cognitive processes. Cambridge University Press, Cambridge/England

Rowlands M (2001) The nature of consciousness. Cambridge University Press, Cambridge

Rowlands M (2002) Two dogmas of consciousness. J Conscious Stud 5–6. Special edition, Noë A (ed) Is the visual world a grand illusion?

Rowlands M (2003) Consciousness: the transcendentalist manifesto. Phenomenol Cogn Sci 2(3):205–221

Rowlands M (2006) Body language: representation in action. MIT Press, Cambridge, MA

Yarbus A (1967) Eye movements and vision. Plenum, New York

Attention in Context

P. Sven Arvidson

Many figures in phenomenology, including founder Edmund Husserl, have uniquely addressed what cognitive scientists call attention. For example, Husserl (1991) discussed dynamic temporal attention, attentional shifting (1970), attentional capture (1982), and serial attention (2001). Jean-Paul Sartre (1956), Maurice Merleau-Ponty (1962), and Aron Gurwitsch (1964, 1966) discussed attention in terms of gestalt principles, including the nature of visual search, illusions, salience, and context effects, often criticizing various psychological concepts of attention. But for each of these phenomenologists the context or margin in attending was the main interest not the focus of attention. In contrast, experimental attention research emphasizes the focus of attention. Things are changing. Phenomenologists still emphasize context and margin, but now cognitive scientists are starting to as well. This means that the theoretical trends in contemporary attention experiments and a century of phenomenology of the structures of consciousness and perception, can both be used to put attention in context. This chapter shows that current cognitive science of attention substantially intersects with a gestalt-phenomenology of attention, even if this intersection is not yet effectively articulated by phenomenologists or utilized by experimenters in formulating hypotheses, models, and theories. It also suggests that phenomenology and cognitive science of attention can be co-revelatory in theory and practice.

I will use a gestalt-phenomenology of attention as the main context within which to connect laboratory experiments with phenomenology. Much of it was originally expressed as a phenomenology of "the field of consciousness" by Aron Gurwitsch (1901–1973). I use Gurwitsch here for three reasons. First, although Gurwitsch would have categorized himself as a Husserlian philosopher of consciousness rather than attention, his work on both can be interpreted in a way that gives the most systematic phenomenology of attention available. Second, his life-work aligns well with cognitive science of attention – he critically interpreted the work of William James, researched brain-injured veterans in Frankfurt in the 1920s (Embree 2004), and interpreted attention in terms of gestalt principles, as a number of cognitive

P. S. Arvidson
Department of Philosophy, Seattle University, Seattle, WA, USA

scientists have begun to do within the last decade (e.g., Lamy et al. 2006; Quinn and Bhatt 2005; Vecera et al. 2004; Roney and Trick 2003; see Palmer 1999 for review). Third, I believe non-phenomenologists will find Gurwitsch's work more accessible (not necessarily more important!) than the other phenomenologists listed above.

My interpretation of Gurwitsch's work is a restructuring of the phenomena involving attention and consciousness in phenomenology. Attention becomes the formative constituent not consciousness. For example, I replace the "field of consciousness" metaphor with the metaphor of a "sphere of attention" (a dynamic, embodied attending in the world, not a mental container or representation) for reasons fully explained elsewhere (Arvidson 2006). These reasons include the argument that the sphere metaphor better acknowledges the depth of dynamic attentional processes than a field and that the sphere metaphor accounts better for the peculiarities of subjectivity in attentional life. It is a sphere of attention rather than a sphere of consciousness because I take attending to be the center of experience, and consciousness to be part of the sphere of attention, as context and margin. Note that what follows is general enough that my unique interpretation of Gurwitsch's work does not play a significant role (for in-depth comparisons of Gurwitsch, Husserl, and Sartre on attention see Arvidson 2006).

A Gestalt-Phenomenology of Attention

On a summer afternoon, I relax while sitting outside the café, thinking about writing, about the friends that are coming soon, watching the people mill around the café. I am attending to these different things at different moments. Then an unusual yellow sports car arrives to parallel park near me, capturing my attention. I watch it park. What have I attended to in this 10 seconds or so? The unique car is the focus of attention, the theme of attention, and it is presented for me in the context of the other cars nearby. The car is presented as a more or less centralized, consolidated gestalt – a thematic whole whose constituents, such as the sleek shape and yellow color and precise movements, have a functional significance for each other (Wertheimer 1921, 52; Gurwitsch 1964, 115). Some constituents of this theme are more prominent and some are less prominent, but none of these constituents of the theme is the focus of attention, the car parking is. The unique car happens to be presented in the context of the surrounding cars, it appears under the light, perspective, and orientation of these background cars. In addition to the gestalt-coherence of the theme, this connection between the theme and context in attending is a second type of gestalt-connection – a unity by relevancy between the theme and its context. The surrounding cars are each gestalts but *not* centralized like the theme. These items, the car in front of the space, the car behind the space, perhaps the nearby parking meter, are more or less clearly presented as relevant to the thematic unique car. The music from a nearby sound system and the sight of a group of customers entering the café happen to be irrelevant to the car as theme, and so are presented as marginal not contextual. That time has passed, that I am sitting, and that I am peripherally aware of myself is

also all marginal at this moment. The margin is largely a horizon in attending, presenting content external to the relevancy that holds between theme and context. As a gateway to this horizon, a marginal halo presents what is related but still irrelevant to the theme. For example, the unseen person driving the car happens to be irrelevant to the theme, but it is related. He or she might become relevant given a change in the theme or its thematic context, such as a growing incompetence in the parking job.

In sum, a gestalt-phenomenology of attention articulates three distinct but related dimensions: *thematic attention* (attention in the dimension of theme or focus), the *context of attention* (consciousness in the dimension of thematic context), and the *margin of attention* (consciousness in the dimension of margin as halo and horizon). *Thematic attention* is centrally focal attention. Thematic attending allows content to become segregated from the thematic context and centralized as a gestalt within this context. The *context* of thematic attention allows content to become consolidated as non-centralized gestalts, relevant to the theme and to other thematic context content (rather than being thematic themselves). The *margin* also allows content to become consolidated as non-centralized gestalts, but segregated from the theme and thematic context, and co-present with them yet irrelevant to them. The attending situation is much more complicated than I have articulated in the parking car example. The point is that each moment of attending life is structured into three dimensions, *attention-in-context-with-margin*.

Achievement of tasks by subjects in numerous attention experiments involves keeping task information central, cuing information relevant, and distracting information irrelevant. The success or failure of the subject to achieve this threefold organization of information in attending is measured in reaction times or through statistical analysis of response accuracy. This tension between maintaining irrelevant information as irrelevant, and targeted information as central, is the tense connection between margin and theme in experiments. Add the maintenance of relevant information and we have affirmation of this three-dimension, gestalt-inspired, organization of attention that phenomenology can articulate. Currently, cognitive science does not propose an adequate theoretical framework for expressing the gestalt-segmentation processes that are marginally and contextually co-present in attending, but are not the focus of attention (cf. LaBerge 2002, 223).

The Context Problem in Attention Research

The good news is that cognitive scientists are now more aware of context than ever because some new areas of research must explicitly acknowledge the role of context in attending. The bad news is that as experimental attention research has blossomed, the operational definition of context has become muddled. A phenomenology of attention can help bring some order to these cognitive models. In phenomenology, cognitive science, and everyday language, attention refers exclusively to the focus, target, goal, or theme. For example, we do not normally say "attention to context" (but see Yang and Lewandowsky 2003, 676; Davenport and Potter 2004, 561, 563) or "attention to the margin" (but see Sartre 1956, 10; and "attention to ground"

in Palmer 1999, 283) even though both of these make perfect sense if each of these dimensions are thought to have a distinct organizational structure from the focus (Arvidson 2004). Yet cognitive science of attention is now venturing beyond the focus in its modeling. The trajectory of attention research is toward conceiving attention as the center of a wider structure or patterning of organized processes, where what is relevant for the focus is distinguished from what is irrelevant, and all three regions have their own determinable organizational processes. In other words, the teleology of attention research is directed toward assuming and examining the three-part phenomenology of attentional life – theme, context, and margin.

Context has always been an issue in attention research. William James (1981) sharply distinguished only focus and margin in 1890, and vaguely pointed to context as "fringes" or "transitive states," (Arvidson 1998), and some influential cognitive scientists still use this vague terminology (e.g., Baars 2003). But since James' work, attention has been only selective attention, the accomplishment of filtering, windowing, spotlighting, channeling, integrating, what is attended from what is *un*attended. In other words, selective attention negatively defines the context and margin in attending. This century-old negative stance toward what lies outside the focus has hampered attention research. For example, a recent gestalt-based, experimental report of contextual organization, calls this organization "grouping without attention" (Lamy et al. 2006). Since the 1980s, these outside regions or factors have been conceived as interfering or facilitating attention by slowing down or speeding reaction times, for instance, spatial cuing (Posner 1980), flanker effects (Eriksen and Schultz 1979; Miller 1991), and contextual cuing effects (Chun and Jiang 1998, 1999). Even so, instead of defining these context and margin regions in terms of their own organizational principles (e.g., in terms of gestalt psychological principles, spatial and temporal principles, etc.), in current scientific research they are most often lumped together without regard to a *positive* account of relevancy. They are "unattended stimuli," "unconscious content," "unselected areas of the field," "irrelevant data," "non-target items," "distractors," or also "filler items" and "noise."

The word "scene" is appearing much more often in attention research in the last 10 years (for an often cited early mention see Biederman 1972), especially in research on tracking of attention (Scholl and Pylyshyn 1999; Pylyshyn 2003; Horowitz et al. 2007), scene perception (Henderson and Hollingworth 1999; Rensink 2000; Davenport and Potter 2004; Becker and Pashler 2005) and scene memory (Zelinsky and Loschky 2005; Hollingworth 2005; Intraub et al. 2006). However "scene," as whatever is co-presented with the theme, is used ambiguously in attention research to mean either a meaningful situation (context) or a mere environment (margin). In an often cited investigation of "serial memory within a scene," Irwin and Zelinsky (2002) tested subjects' ability to visually remember a "scene" operationally defined as seven objects (such as a teddy bear) in one of seven fixed locations in a baby's crib displayed on a video screen. "Scene perception" in this experiment was attending to and remembering the items in the crib, the location of each item in relation to the others, prior to the disappearance of the display. The scene was defined by the relation of the items to each other. Amazingly, the definition of scene in this and in the updated experiments (Zelinsky and Loschky 2005, 689n), only incidentally included the *scene* of the items – the crib! In another influential

study, Henderson and Hollingworth (2003) presented subjects with what they call "complex real world scenes" in a number of experiments concerning attention and visual memory. The "complex real world scenes" were briefly presented pictures on a computer monitor, which makes for a very odd "real world." As with Irwin and Zelinsky (2002), these researchers were not interested in perception of the scene *as* context in relation to the focus of attention. They were interested in the serial attention of items within the scene, where the latter is treated as a mere environment, not a context. How is this scene perception? Hollingworth (2005) makes conclusions about visual memory for "natural scenes" (408) but in the experiments uses "images of natural scenes" (406). Perhaps this is just a matter of brevity in language, but the substitution of laboratory for real or natural scene comes too easily in this kind of research. More importantly, in defining visual short term memory (VSTM) and visual long term memory (VLTM), Hollingworth (2005) pays no notice to *relevancy* in storage or recall, such that information in VSTM or VLTM can be either relevant or irrelevant to what is attended – a notable ambiguity. The various ambiguities in the use of "scene" have recently been noticed by some cognitive scientists (Bülthoff and van Veen 2001; Simons et al. 2002; Ruddle and Lessels 2006; Waszak et al. 2005; Bravo and Farid 2006; see also use of "outdoor scenes" in Frey et al. 2007, and in Lappin et al. 2006).

Though experimenters naturally use the word "scene" because the overwhelming majority of research is on visual attention (Gottesman and Intraub 2002), the more general term for what they mean is "context." There is very little agreement about what "context" means. Context can properly refer to what is materially relevant to the theme, such as the use of context in many narrative comprehension experiments (McNamara and McDaniel 2004; Kelter et al. 2004) or lexical memory studies (Hicks et al. 2005). Unfortunately, context often means the presence of "filler items" (Tremblay et al. 2005; Jescheniak et al. 2005), "distracting contextual elements" (Libera and Chelazzi 2006), irrelevant memory (Marsh et al. 2001), or vague arousal or awareness (Cohen 1993; Coull 1998; Koch and Crick 1994). As with the critique of computer displays as real world scenes, some cognitive scientists are also concerned about what counts as context. In a section refreshingly titled "What is Context?" Malmberg and Shiffrin (2005, 330) write, "Because context is not usually manipulated in an experiment, the construct often serves as a grab bag for all information other than the item information that is varied in the study. This state of affairs can lead to a very broad and fuzzy definition of context. Better and more specific kinds of context, including classification of the various kinds of context information into categories, require embedding the concept in a reasonably rich theoretical framework." One of the points of this chapter is that phenomenology offers this framework.

Connecting Context to Focus

This section connects context in phenomenology of attention with the focus in cognitive science. But first I must briefly articulate the connection between the focus and margin in attending. Marginal consciousness is the process whereby

distracting or neutral information is presented outside the focus of attention as not relevant to that focus. All attention experiments must assume a marginal dimension in attending. Typical language in cognitive science studies that identify the margin are "task-irrelevant processing," "unattended stimuli," "irrelevant information," "distractors," and so on. Negatively, the margin is the dimension of *ir*relevance for the theme. Positively, it is intimately *connected* to the theme: it presents content simultaneous with the theme though irrelevant to it and often in dynamic tension with the theme (cf. Gurwitsch 1964, 282–283; 1966, 272). The margin is a dimension with its own organizational principles and depth (Gurwitsch 1985; Arvidson 2006). Any experiment on attentional capture, spatial cuing, temporal cuing, inhibition, negative priming, multiple object tracking, attentional blink and so on, must assume this distinction and connection between a focus of attention and organized periphery (margin). For example, in cuing paradigms (Posner 1980) an invalid cue for target location is marginally presented at the moment it transforms from valid to invalid (Arvidson 1998, 2003). The cue has transformed from being relevant to the appearance of the target (expectancy) to irrelevant to it (merely co-present). Neurological evidence also points to co-present cortical activity related to the processing of "task-irrelevant stimuli" in attending, according to fMRI and MEG (Downing, Liu, and Kanwisher 2001, 1336). Also, research on memory and attention assumes the theme and margin connection, or else there could be no probing for recall of irrelevant content processed at the time of "encoding" (e.g., Troyer and Craik 2000; Marsh et al. 2001).

Although attention research must assume a marginal dimension in attending, it is important not to conflate the two organizational dimensions. Marginal processing in attending has similarities with focal attention, but the differences are profound and conclusive. One similarity is that both dimensions present gestalts, one as potential themes and the other as a current theme. Gestalts of any type are comprised of constituents that have at least an emerging functional coherence for each other. Gestalts may be presented in any mode of achievement, from relatively stable, well-formed gestalts to partial, dim, diffuse, or dynamic gestalts, unstable or not well-formed but possibly proceeding toward it. The marginally presented gestalts are not unified with the theme in any material way, unlike gestalts in the thematic context. These marginal gestalts have an "and-connection" to the thematic gestalt and contextual gestalts (Gurwitsch 1966, 233–234; Koffka 1925).

A theme is always presented within a thematic context (Gurwitsch 1966, 203; 1964, 319). Context in attending involves the presentation of information or items relevant to the theme, and these items are organized by a unity of relevance for each other and for the theme. A well-developed thematic context has a gradient from near to remote relevance with respect to the theme, with information more intensely relevant nearer to the theme, and information less relevant (but not merely related or irrelevant) more remote. Although the theme appears in the light, perspective, and orientation (a positional index) bestowed upon it by the thematic context (Gurwitsch 1964, 360), the theme dictates the lines of relevancy for the thematic context (Gurwitsch 1966, 204 and 207).

Two leading attention researchers, Chun and Jiang (1998, 1999), recently developed a new experimental paradigm – contextual cuing – that explicitly calls forth the function of a thematic context in attending. "Contextual cuing refers to improved performance in visual search tasks based on learning association between targets and surrounding visual context" (Chun and Jiang 2003, 224). In a typical contextual cuing experiment, subjects search a computer display screen for a rotated letter T among rotated letter L's (these latter are called "distractors"). In each unique display, the configuration of the L's cued the location of the target T. After a certain number of trials, subjects were better able to find the target T among the L's. They learned the relevant context for where the T would appear in various, unique configurations. The contextual relevancy here is not that T's are related to L's, as letters to letters, in the way that a family member is singled out in viewing a family photo. It is only a spatial relevancy: "Visual context can be defined as the global configuration of all the items" (Chun and Jiang 2003, 224). This is fine, since these are spatial visual search experiments. But these and other researchers do not have a model of context that includes more diverse relational characteristics, or the relevant environmental context (for more on the latter see Lappin et al. 2006). "Contextual refers to the impact of other information, typically co-occurring items, on the processing of the target" (Jiang and Chun 2003, 278). Although spatial contextual cuing experiments wonderfully affirm the *existence* of a contextual dimension in attending, in hypotheses and discussions of results this contextual dimension is not distinguished as a unity of relevancy for the theme, and is not distinguished from the irrelevant margin.

Like spatial contextual cuing, temporal contextual cuing must assume and affirm a contextual dimension in attending. In a typical temporal cuing experiment (e.g., Olson and Chun 2001), a subject is presented with a stream of visual images on a computer monitor and asked to complete a task. Implicitly, through the rhythm of the streaming, the subject develops temporal cues that facilitate attention to the target. This is a "pattern-based expectancy" that views context in terms of relationships among features, a temporal context of rate and rhythm, for example, in attending to ongoing speech or music (Barnes and Jones 2000, 261). Recently, Jones and McAuley (2005) have updated previous work (McAuley and Jones 2003) within the temporal contextual cuing paradigm. The significant conclusion is that they affirmed three levels of context that contribute to time judgments: global context (pace of session), trial-to-trial context, and local context (intra-trial pace). That is, they are hypothesizing about and modeling the context in attending, differentiating various organizations within it. Jones and McAuley (2005) not only found that there are near (local) and remote (global) portions of the thematic context, as also claimed by Gurwitsch (1964, 338, 353, 379; 1966, 205; see also Olson and Chun 2001, 1309), but they found that the local and global contexts were relevant to each other, with the local context more primary in the "entrainment of tempo" (2005, 412). In other words, these experiments not only take seriously the organization of the context itself, they support a phenomenology of attention prediction that even the more remote regions of the thematic context must nonetheless be relevant to those more near the theme (Gurwitsch 1966, 206), within an indefinite continuation

of context (Gurwitsch 1964, 379). This remote zone of the thematic context is experienced as unlimited capacity through pointing references (Gurwitsch 1964, 380; Intraub et al. 2006), while the near zone appears to be limited capacity (Scholl 2001; Noles et al. 2005).

The context in attending is codified in the language of diverse research concerns, such as attentional control, semantic priming, narrative comprehension and frames of reference. In research on the control of attention, which is concerned with bottom-up versus top-down processes, the context in attending is often referred to as "attentional set" or "attentional control setting" (Folk et al. 1992). Attentional setting generally refers to preparatory or co-present emotional or attitudinal orientations that affect or frame focal attention, such as experimenter instructions to subjects (see also Simons and Mitroff 2001). In attentional control studies, context is connected to focus through expectation. In semantic priming experiments, in which the processing of a target word (e.g., DAUGHTER) is facilitated by a prime word (e.g., MOTHER), researchers call this contextualization of the target an "activation of the semantic priming system" (Smith et al. 2001, 1289) or "activation of relevant knowledge" (McNamara and McDaniel 2004, 479) or "resonant connections" (Nelson et al. 2003). In these studies, context is connected to focus through salient semantic knowledge. In the narrative comprehension paradigms, which study the relation between the ongoing situation of a protagonist in a text and new information, thematic context is codified as a mental representation in a "situation model" (Zwaan and Radvansky 1998; Zwaan and Madden 2004; cf. Kelter et al. 2004). In these studies, context is connected to focus through situation updating in narrative. In frames of reference paradigms, which study spatial perceptions and actions in space, contextualization involves relative contributions of reference systems – ego-centric (e.g., body coordinates) and environmental (e.g., walls, latitudes) (Mou et al. 2004; Lappin et al. 2006). In these studies, context is connected to focus through perceived distance or spatial recall from narrative.

One of the characteristics of the gestalt-connection between the theme and thematic context is that the context bestows upon the theme an orientation, light, and perspective – a *positional index* (Gurwitsch 1964, 362). For example, in experiments on temporal contextual effects, the targeted theme is attended to as an event in a series of events, and has a unique and distinct position within that context. Speaking loosely, we might say that the theme derives its meaning and significance from the context. Operationally defining positional index advances beyond the mere fact of context effects and investigates the organization peculiar to the context itself, as in the tri-partite temporal context experiments (Jones and McAuley 2005) discussed above. In a fascinating examination of positional index (they do not call it that), Roney and Trick (2003) manipulated how context affects the gambler's fallacy. "The gambler's fallacy is the tendency to see a given outcome as less likely if it has just repeatedly occurred, in this case, leading to the choice of tails following three heads. It is a fallacy to the extent that the person's expectancy deviates from the true probability of getting heads in a coin toss (50%)" (Roney and Trick 2003, 69). A subject presented with four tosses, in which three were heads, would predict the emergence of the upcoming theme (the fourth toss) according to this improbable

context of heads. Given the context, subjects predicted a higher probability that the fourth toss would be a tail. The context bestowed a certain character or orientation upon the theme. When the experimenters ended a block at three head tosses, for example, and started a new block with the next toss, which was still the fourth toss out of three plus one, subjects did not tend to predict a higher probability that the fourth toss would be tails. In other words, solely as a matter of context replacement, the orientation of the theme had substantially changed, a result predicted by Gurwitsch's concept of positional index. Gamblers recognized the difference between continuing with an existing series of gambling, or *beginning* to gamble in a new series. Similar support for these rich context effects of positional index occur in a variety of experiments, for example, in aesthetic judgment (Russell 2000), judgment involving quantifiers (Newstead and Coventry 2000), body and imagery (Hanrahan and Vergeer 2001), emotion contextualization (Kunzendorf et al. 2000), mortality salience (Arndt et al. 2002), decision making (Sharps and Martin 2002), lexical processing (Kambe et al. 2001; Rawson and Kintsch 2002), proprioception (Simons et al. 2002), counterfactual thinking and emotions (Mandel 2003), mood maintenance (Watkins et al. 2003), viewer frames of reference (Mou et al. 2004; Waszak et al. 2005), visual working memory (Olson and Marshuetz 2005), and object-background semantic consistency (Davenport and Potter 2004).

Memorial indexing and visual indexing appear to be positional indexing in Gurwitsch's sense. For example, OSCAR (OSCillator-based Associative Recall) emphasizes temporal contextual uniqueness in serial order memory (Brown et al. 1999). The general idea of OSCAR is that the item is positioned within an overall dynamic context with other items, as thematic attention serially shifts from item to item in the list, a positional indexing in which the theme gets its orientation and character from the context, which in cases of recall includes previously attended serial items which are now presented contextually. Recently, a number of experiments have investigated "visual indexing" – a "pre-attentive process" that assigns a positional index to an object outside of the attentive focus (Pylyshyn 2003, 2001; Scholl et al. 2001). A problem these and other researchers are trying to solve is how subjects can keep track of a number of objects that are not currently focal, and are possibly changing positions in the "field of vision." This multiple object tracking (Sears and Pylyshyn 2000; Horowitz et al. 2007) is a problem of positional index, of orientation of the theme in the thematic context.

Achieving the Bigger Picture in Cognitive Science of Attention: Attention-in-Context-with-Margin

A recent editorial in *Psychological Research* acknowledges that, "In some way, people must be able to configure and re-configure their cognitive system in a way that task-relevant information is picked up, maintained and stored efficiently, and that appropriate actions are prepared, planned, and then executed in the light of the available information. But we are only beginning to understand how this configuration works"

(Hommel et al. 2002, 215). I will sample a few researchers who are beginning to distinguish and interrelate focus, relevancy, and irrelevancy (theme, context, and margin). Like all the other cognitive scientists discussed above, these authors do not yet see their models and hypotheses in light of the three-dimension phenomenology of attention.

Researchers who worry about ecological validity (Bülthoff and van Veen 2001), embodied attending (Waszak et al. 2005), navigation (Ruddle and Lessels 2006), and environmental factors (Holland et al. 2005) naturally appeal to context and margin in attending. Bülthoff and van Veen (2001) use virtual reality to bring more ecological validity to object-recognition, visual scene analysis and navigation. For example, to model urban bike-riding navigation and recognition, they use special panoramic screens and feedback loops between perceptions, movements, and stimulus. The result is that attending is studied "in a realistic context," in which "In essence, we have greatly improved stimulus relevance" (p. 245). Putting subjects in a virtual environment acknowledges the margin in attending in a positive way, namely, as the co-present horizon of body and world, and the "dynamic scene analysis" these experimenters perform is an analysis of relevant context (also see immersive virtual environment and embodiment in Ruddle and Lessels 2006).

Fortunately, there are a number of other attention and memory studies that could be interpreted similarly. Braver et al. (2001) examined "context processing" in older adults. Employing connectionist computer-based modeling, they use the term context as the foundation for *everything* in attending (cf. Gurwitsch 1964, 358), for what I have distinguished as theme, context, and margin. So the first point is that they allow context to be broadly inclusive and relevant, "Thus, context representations may include a specific prior stimulus or the result of processing a sequence of stimuli, as well as task instructions or a particular intended action" (Braver et al. 2001, 747). The second point is that these experimenters distinguish three domains or functions in attending – attention, active memory, and inhibition. Although "context" underlies all three domains, it functions differently in each, and the result is that the three domains these authors describe can roughly coordinate with theme, context, and margin. In addition, the "context processing" they describe easily fits the articulation of gestalt-connections in a phenomenology of attention: the gestalt-connection of *unity by relevancy* among items in the thematic context, and *between* the theme and the thematic context, and the connection of *irrelevance* between the margin on the one hand and the theme and its thematic context on the other. The connection of irrelevance implies some middle process, which I would call the halo, which for these authors is a gating mechanism (Arvidson 2006). Braver et al. would likely be cautious about my way of putting all of this. My main point is that their experiments and modeling of the results support the notion of attention-in-context-with-margin. Also noteworthy is that Braver et al. (2001, 757) define context and working memory separately, since working memory involves identification, whereas "contextual processing" may not (see also Engle et al. 1999).

There are more examples of positive articulations or explicit assumptions of attention-in-context-with-margin. In reality, all attention research involves the three dimensions in attending. For instance, the many negative priming experiments in the literature can only be designed by assuming three dimensions in attending

(Arvidson 2003). Maintaining the narrow focus on the focus of attention in hypotheses and modeling is fast becoming the old way of doing attention research, and the new way brings cognitive science and phenomenology into better dialogue.

Dynamic Attention: Context Transformations, Theme Replacements, Attentional Capture

The sphere of attention is a dynamic tension. Instead of discussing organizational principles within each dimension in attending, this section articulates common transformations between the dimensions (for a fuller account, see Arvidson 2006). In what follows, I have adopted and expanded what Gurwitsch calls "thematic modifications" (1966, 223–267), and what cognitive scientists often call "attentional shifts." The transformations are typical and regulated: typical because they are distinguishable in direction and locality (e.g., theme or context or both), and regulated because a gestalt or gestalt-connection may admit transformations of specific types (Gurwitsch 1966, 223; see Husserl 1970, 166–167). Particular transformations are not inevitable, each case is contingent upon saliency and external conditions (see Gurwitsch 1964, 103). Any particular event in attending likely involves some combination of these transformations. As is the case throughout this chapter, although I use the language of a phenomenology of attention when I interpret and report what experimenters are doing, they do not use this language themselves.

Context Transformations

In all of these context transformations, *enlargement*, *contraction*, *elucidation*, *obscuration*, *context replacement*, the theme endures while the context changes. There may be more than these five kinds.

Enlargement. An enlargement of the thematic context means the significance of the context for the theme grows while the theme itself remains the same (Gurwitsch 1966, 223–227; Husserl 1982, §67). Some attention researchers call this process "zooming out" (Eriksen and St. James 1986; Pasto and Burack 2002), which is too simplistic (Arvidson 2000; Gurwitsch 1966, 226). Enlargement is at work when a teacher asks a student to expand the relevance of a paper thesis, or when we ask a friend to put a recent tragedy in larger perspective. Enlargement may be also be essentially involved in what is called "joint attention," when young children begin to add another's attending to the context for the thematically presented toy (Rochat 1999; Arvidson 2003). Enlargement and contraction work in opposite directions.

Contraction. A contraction of the thematic context involves a narrowing in significance of context for theme (Gurwitsch 1966, 224). Contraction may be the key attentional transformation in depression (Jacobson et al. 1996), which involves a

contracted relevancy of the future (MacLeod and Salaminiou 2001), or in repressors who contract context when information becomes too personal (Mendolia et al. 1996; Terry and Burns 2001). Fatigued obsession in thematic attention to a problem of some sort can be accompanied with contraction of the thematic context for the theme (see Rees and Lavie 2001; Gurwitsch 1964, 336). Contraction of context is likely involved in navigation and tasks, especially in a variable environment: as activity moves from planning to control, what is relevant for the action shrinks (Glover and Dixon 2001); also, as expertise increases, the range of concerns decreases (Lewandowsky and Kirsner 2000).

Elucidation. An elucidation is a clearing of an obscurity in the thematic context. Unlike enlargement, which increases content not previously presented through a broadening of relevance, elucidation clarifies what is already presented contextually (Gurwitsch 1966, 224–225). Since the context is always somewhat obscure, elucidation is never completely successful (Gurwitsch 1966, 226). Elucidation may be primarily involved in some Buddhist-based mindfulness-awareness practices, where attending includes a clearing up of context (Varela et al. 1991, 79; Haywood 1998, 613; Wallace 1999, 177). Some insightful problem solving appears to involve elucidation (MacGregor et al. 2001). Experimenters have found that adding context to a pun in a way that elucidates the context leads to greater perceived humor (Lippman et al. 2001), while others used context elucidation to help possible victims reconstruct the setting of a crime (Hershkowitz et al. 2001), which is not primarily an addition of new information (i.e., enlargement) but a clarification of existing relevancies. In examination of decision-making, Sharps and Martin (2002, 274) found that providing subjects with "relevant information on the immediate context of the decision," rather than leaving it obscure, elucidated the context for decision making and significantly helped subjects recognize negative quality decisions. By manipulating "background information," Rawson and Kintsch (2002) found that an elucidation of the context improved memory for text content in reading – that is, elucidation of context sharpens positional indexing of thematic content, a result with very practical applications.

Obscuration. Obscuration conceals some relevance of the thematic context for the theme. A clarified context becomes less defined, general lines of relevance transform from many to few. Obscuration is not contraction because the breadth of the context is not significantly reduced in obscuration, just the strength of lines of connection. For example, bizarreness can affect memory, leading to false memories, by disrupting existing gestalt-connections between contextual items, obscuring relevancies (Worthen and Wood 2001). Obscuration, probably in concert with contraction, is involved in what psychoanalysis calls repression. When someone tells me something about myself that I would like to hide, I can at first clearly understand the truth of what they say, and then dim that truth through obscuration (and/or through contraction), the clear relevancies are made nebulous. Also, since psychoanalysis assumes the need for elucidation of obscured memories, symbols, and so on, such as in dream or childhood experience, where the obscuration is taken as a symptom or problem, attention research may connect with psychoanalysis on this score.

Context replacement. Context replacement can be a more radical modification in attending than the others in this grouping, since it is a replacement not a development of the same context (Gurwitsch 1964, 322). The above account of the gambler's fallacy in the Roney and Trick (2003) experiments involved a context replacement. In the fourth toss, the theme itself did not change, but the perspective or orientation it derives from the new context changes (Gurwitsch 1964, 359). It is obvious that changing the perspective through context replacement from "do I want to continue to gamble" to "do I want to start gambling" could be critical for gambling therapy, but context replacement could also be crucial for a nurse's hemophobia or a professor's glossophobia. In examining cognitive change in psychotherapy, Brewin (1989, 388) writes "the therapeutic task is to recreate a plausible context that will enable patients to encode subsequent experiences in a more discriminating way, rather than perceiving them as all indicating failure, rejection and so on." Similarly, Lewandowsky and Kirsner (2000) describe how the inability of expert Australian bush fire controllers to replace the content of the context with a new one is linked to error (see also Botvinick and Plaut 2002). The concept of context replacement is also used in lexical and narrative comprehension research, where a context replacement can affect semantic ambiguity (Jescheniak et al. 2005) or help account for a dynamic view of comprehension (a "fresh start" in Kelter et al. 2004).

Theme Replacements

I describe four kinds of theme shifts, *serial shifting, restructuring, singling out, and synthesis*. In all of these transformations the theme is replaced with a new theme.

Serial-shifting. Serial-shifting occurs when the theme is replaced by a new theme that is relevant to it (Gurwitsch 1964, 345; 1966, 230–232; Husserl 2001, 292). Essentially, the relevant context for the old theme provides the item that will become the new theme, such as when one is engaged in step-by-step procedural tasks or problem solving. The temporal context experiments discussed above were examples of serial-shifting research (e.g., Olson and Chun 2001), as are the relatedness effect or the semantic priming effect (Rafal and Henik 1994; Smith et al. 2001; see Arvidson 2003). Since serial-shifting, accompanied by singling-out (selection), has long been considered the standard accomplishment of attention, it is well researched. Also, it is the standard conceptualization of attention in memory research because so many experiments have been designed around the recall of serial ordered lists.

Restructuring. Restructuring, singling out, and synthesis each involve a substantial change in the configuration of the theme itself (Gurwitsch 1966, 237–248). Cognitive scientists of attention rarely recognize that these transformations involve the presentation of a *new* theme with a *new* thematic context in the course of achievement (on achievement see Gurwitsch 1964, 103; see also Husserl 2001, 57). This achievement is a replacement, not a shift. Multi-stable figures, such as vase-faces

or Necker cube line drawings, are standard examples of restructuring. Restructuring is a substantial change in the function of the formative constituents of the theme. Formative constituents are dominant in the presentation of the theme, while formed constituents depend on these. In a crowd of people, some are presented as primary, even though the crowd as a whole is the theme. For example, as the unique sports car parks in the space near the café, the formative constituents, perhaps the car's sleek shape and yellow color and precise movements, characterized the car parking in the space, while the space is presented as a dependent (formed) constituent of the theme in a functional significance. However, this theme could restructure so that the formative constituents become formed, and the previously formed become formative: what is presented as thematic is not now the car that is parking in the space, but the space that the car is parking in. The transformation is confined to the thematic dimension, but it is not a serial-shifting, singling out or synthesis, and it is not primarily a context shift, although it may include one. In the vase-faces figure, for example, restructuring involves the transformation of a formative constituent of a theme (e.g., the lips in the faces figure) into a formed constituent (e.g., the ornamental protrusion on the stem in the vase figure) in the presentation of a *new* theme, in this case the vase (Gurwitsch 1966, 237–240 and 14; also 1964, 118–119). Varela (1999) has suggested that restructuring is much more common than is usually considered, and I would agree. For example, a study in the delay of gratification in children induced psychological distance to desire for a cookie by restructuring the theme in favor of shape and color rather than taste (Mischel et al. 1989). In our goal-oriented lives, how many times a day or week do we stay on task through similar attentional restructuring strategies?

Singling out. Singling out is when a constituent of a theme is attended to thematically, so that this constituent becomes a theme itself (Gurwitsch 1966, 240–243; Husserl 2001, 295 and 298). Singling out falls under the generic phrase "attentional selection," or just selection, in the attention literature. Alas so do serial-shifting, synthesis, and restructuring. An analysis of dynamic attending must distinguish these better. Singling out is sometimes called "zooming in" (Metzinger 2003) and explained using suspect metaphors of attention (Arvidson 1996; Gurwitsch 1966, 265–267). Singling out is a replacement of one theme with another. For example, singling out the shiny, chrome wheel of the sports car so that it is the focus of attention does not leave everything else the same. The car that is parking is not the theme any longer, it has been replaced, and the focal wheel likely brings its own new context, such as the car as a whole, but now this car is in the distinct dimension of context (see Gurwitsch 1966, 243). Unfortunately, the transformation involved in singling out is often glossed-over or goes unrecognized by cognitive psychologists (Arvidson 2003), for example, in the global-local manipulations using the Navon letter (Navon 1977; Rauschenberger and Yantis 2001; Fileteo et al. 2001).

Synthesis. Synthesis is the transformation of a theme into a constituent of a new larger theme (Gurwitsch 1966, 243–248; Husserl 2001, 176 and 206), as when the parking sports car becomes a constituent in a larger theme of the row of cars, or a previously thematic person becomes a constituent in a group of people. The sports car or the person is no longer a theme. This is a more radical transformation than

recognized in attention literature because a new theme is presented, with new internal relations, and with new relations to a possibly new thematic context. Synthesis is often called a "spreading of attention" in object-based attention research (Scholl et al. 2001, 172; Marino and Scholl 2005, 1141; Chong and Treisman 2005). This is a vague term in light of the transformations under discussion (see Gibson and Kelsey 1998, for a critique of "spread" of attention). A more useful account of synthesis comes from Kelter et al. (2004). They recommend a dynamic view of narrative comprehension and propose an updating process called "tracking." Comprehenders take each new event, as long as it is a situational continuity sentence ("Jane was lying in bed. *She turned off the stereo*.") as a constituent in a new theme. This synthesis is a description of a dynamic now in comprehension. Also noteworthy, Hollingworth et al. (2005) recently demonstrated the process of synthesis ("figural grouping") in visual short-term memory. Successful teaching and learning can involve synthesis and singling out (Clark 1999, 188), and synthesis is involved in intuition or insight (Bastick 1982; Arvidson 1997), and likely has a significant role in creativity (Raidl and Lubart 2001; Williamon et al. 2002).

Attentional Capture

As the sports car pulls in to park, let us say the shiny, chrome wheel especially captures my attention and is singled out. This capture is similar but distinct from a sudden crash of plates and glasses nearby that may capture my attention. The latter is a margin to theme attentional capture, rather than a singling out, and is most common in the literature, and hence it is what I will mean by attentional capture here. Attentional capture is an orienting of attention toward a new location or object even when one is highly engaged in some other project or pursuit, that is, even when attention is elsewhere or otherwise oriented (Gurwitsch 1966, 272; Husserl 2001, 127 and 197). Since attentional capture is an identifiable onset and offset of activity, it is an important phenomenal link between cognitive science models of attention and neurological models (Posner 1995; Kanwisher 2001; see Finkbeiner et al. 2007 for a recent example). Attentional capture research is about control. Some argue that shifting attention to a location or object is primarily bottom-up or involuntary (Theeuwes 1991, 1994, 2004; Yantis 1993; Theeuwes and Chen 2005). Others produce evidence that shifting attention to a location or object is mostly or originally controlled, top-down, endogenous, voluntary (Folk et al. 1992; Watson and Humphreys 1997, 2005; Folk and Remington 1998; Dalton and Lavie 2004). Aron Gurwitsch (1964) instructively anticipates this controversy.

Gurwitsch argues that saliency, not selectivity, underlies attentional capture and attentional control in general. Elsewhere (Arvidson 1992, 2008) I have discussed Gurwitsch's argument in some detail so I will only point out some main themes here. Like cognitive scientists, Gurwitsch (1964, 99) argues that learning is a relatively permanent modification of the psycho-physical system (a change in what he calls "internal conditions"). This change, along with prevailing "external conditions," means that

certain content is more likely to become segregated and emerge in attention (Peterson and Enns 2005). Selectivity cannot account for this stabilization of processing, while saliency can. An achievement in attending, such as the dynamic bringing into focus of the plates and glasses crashing, is a replacement of what was previously thematic (e.g., the car parking). The crashing is almost immediately attended as meaningful, and the "almost" refers to the transformation in attending between what was thematic and what is now becoming thematic (perhaps a function of the so-called "attentional blink," see Arvidson 2003). Even if an "attentional set" is prepared as I search for something hidden in dense clutter, such as an apparently tardy Pierre in a Paris café (Sartre 1956, 10), when Pierre emerges as thematic, the previous evanescent theme of Pierre *as* absent is replaced by a new theme, Pierre *as* present (on search in dense clutter, see Bravo and Farid 2006). If I am instructed to attend to the teddy bear in the crib, a top-down, endogenous "control set," the attending nonetheless ends up with the *emergence* of the teddy bear in its context. In short, order and organization are not bestowed upon the theme, the gestalt-coherent organized theme emerges as independent of and unified with a relevant context. Willing is merely preparatory to this attending, and in fact, is attending to something else. Endogenous selective attention yields to the saliency of exogenous "selection" (Arvidson 2006), establishing over time a kind of facilitation or inhibition of any particular endogenously initiated selective attention (Gurwitsch 1964, 103; see Ohman et al. 2001, 466–467 on "learning history"; and Cavanagh et al. 2001 on "sprites"). I have come to know what a café is, what glasses and plates are and how they can crash, and so on, such that given certain external conditions, they saliently emerge in attending.

The point of this section is that attention is a dynamic embodied attending in the world, and can be articulated as a dynamic tension between contents within and between the three dimensions, theme, context, and margin. There is a constant dynamism and tension in attention (Arvidson 2000). For example, the restructuring from faces to vase is a movement in the theme, a dynamism that is not distinct from the two presentations. There is also a tension in attention, a push and pull between organized content, such as between two views of multi-stable objects (Kelso et al. 1995). In attentional capture there is a tension between the marginalization of the present theme and the encroaching marginal gestalt. I believe cognitive scientists are starting to recognize that attending activity is a process of transformations more or less dynamic and tense at each moment (e.g., see "visual system reconfiguration," Di Lollo et al. 2005; language governed "attention shifting", Taube-Schiff and Segalowitz 2005; Morgan and Meyer 2005; "tracking" in narrative comprehension, Kelter et al. 2004; "temporal context," Barnes and Jones 2000).

Conclusion

What can cognitive science of attention learn from phenomenology? First, attention researchers frequently look at context (and margin) with one eye closed. Many of them model context in implicit ways or use the concept ambiguously

within and across very similar paradigms, for example, spatial cuing and attentional capture. Why are so few asking aloud, "What is context?" (Malmberg and Shiffrin 2005). Opening the other eye will lead to granting context its proper, unambiguous status in attending, giving it genuine, constant presence along with focus. This explicit acknowledgment that context is a positive part of attending will make findings more relevant and widen the significance of attention experiments for memory, learning, perception, and so on. Second, cognitive science should take a more molar view of attention rather than a molecular view. In the same way that phenomenology stresses the lived-world within which it articulates attention, memory, imagination, perception, and so on, cognitive science can be inspired by phenomenology to unhook itself from molecular focus on milliseconds, saccades, and blinks in the focus of attention. These are important, but additional measurable units in attending can be found outside the focus, as discussed above, in the uniquely organized dimensions of context and margin.

What can phenomenology learn from cognitive science of attention? First, phenomenologists should start to think about intentionality in terms of attentionality (Arvidson 2006). Attention is clearly assumed or explicitly invoked in any full sentence about the essence of consciousness. By the same token, the reason why there is so much attention research published in psychology is because attention is central in human life. Taking attentionality seriously will bring phenomenology into more useful contact with this rich, experimental literature. Second, phenomenologists should allow their insights to be sometimes guided by empirical results, and drop the old dictum that its results are tainted since psychology aspires to be a natural science. Exciting journals (e.g., *Journal of Phenomenological Psychology*, *Human Studies*, *Journal of Consciousness Studies*, *Phenomenology and the Cognitive Sciences*) generally advocate for an open discussion of the relation of natural science and phenomenology, and I believe we are in a new era of understanding, even with others pushing back in various ways (e.g., Bruzina 2004; Dennett 2007). One example of how phenomenology of attention can learn from cognitive science involves work on near regions (Noles et al. 2005; Bian et al. 2006) and remote regions (Intraub et al. 2006) of the context in attending (Rensink 2000; Olson and Chun 2001; Pylyshyn 2003; Jones and McAuley 2005; Malmberg and Shiffrin 2005), while another involves the nature of the marginal halo, although researchers do not call it this (Braver et al. 2001; Oberauer 2002; Yang and Lewandowsky 2003; McNamara and McDaniel 2004).

Finally, I advocate for *interdisciplinary attentional studies* along the lines of interdisciplinary consciousness studies that has blossomed over the last 20 years. Rather than the prevailing view that attention is something that happens in the field of consciousness, consciousness becomes part of the sphere of attention – attending becomes the unifying center for discussions of consciousness. A primary advantage is that attention is already deeply researched in cognitive science, and so the squishy concept of consciousness gains some more definite organizational form and function as context and margin in attending.

References

Arndt J, Cook A, Greenberg J (2002) Mortality salience and the spreading activation of world-view-relevant constructs: exploring the cognitive architecture of terror management. J Exp Psychol Gen 131:307–324

Arvidson PS (1992) On the origins of organization in consciousness. J Br Soc Phenomenol 23:53–65

Arvidson PS (1996) Toward a phenomenology of attention. Human Studies 19:71-84

Arvidson PS (1997) Looking intuit: A phenomenological analysis of intuition and attention. In: Robbie D-F, Sven AP (eds) Intuition: the inside story: interdisciplinary perspectives. Routledge, New York, pp 39–56

Arvidson PS (1998) Bringing context into focus: parallels in the psychology of attention and the philosophy of science. J Phenomenol Psychol 29:50–91

Arvidson PS (2000) Transformations in consciousness: continuity, the self and marginal consciousness. J Conscious Stud 7:3–26

Arvidson PS (2003) A lexicon of attention: from cognitive science to phenomenology. Phenomenol Cogn Sci 2:99–131

Arvidson PS (2004) Experimental evidence for three dimensions of attention. In: Lester E (ed) Gurwitsch's relevancy for cognitive science. Springer, Dordrecht

Arvidson PS (2006) The sphere of attention: context and margin. Springer, Dordrecht

Arvidson PS (2008) Attentional capture and attentional character. J Phenomenol Cogn Sci 7:539-562

Baars BJ (2003) How brain reveals mind: neural studies support the fundamental role of conscious experience. In: Anthony J, Andreas R (eds) Trusting the subject. Imprint Academic, Exeter, UK

Barnes R, Jones MR (2000) Expectancy, attention, and time. Cogn Psychol 41:254–311

Bastick T (1982) Intuition: how we think and act. Wiley, New York

Becker MW, Pashler H (2005) Awareness of the continuously visible: information acquisition during preview. Percept Psychophys 67:1391–1403

Bian Z, Braunstein ML, Andersen GJ (2006) The ground dominance effect in the perception of relative distance in 3-D scenes is mainly due to characteristics of the ground surface. Percept Psychophys 68:1297–1309

Biederman I (1972) Perceiving real world scenes. Science 177:77–80

Botvinick M, Plaut DC (2002) Representing task context: proposals based on a connectionist model of action. Psychol Res 66:298–311

Braver T, Barch DM, Keys BA, Carter CS, Cohen JD, Kaye JA, Janowsky JS, Taylor SF, Yesavage JA, Mumenthaler MS, Jagust WJ, Reed BR (2001) Context processing in older adults: evidence for a theory relating cognitive control to neurobiology in healthy aging. J Exp Psychol Gen 130:746–763

Bravo MJ, Farid H (2006) Object recognition in dense clutter. Percept Psychophys 68:911–918

Brewin CR (1989) Cognitive change processes in psychotherapy. Psychol Rev 96:379–394

Brown G, Vousden JI, McCormack T (1999) The development of memory for serial order: a temporal-contextual distinctiveness model. Int J Psychol 34:389–402

Bruzina R (2004) Phenomenology and cognitive science: moving beyond the paradigms. Husserl Stud 20:43–88

Bülthoff H, van Veen H (2001) Vision and action in virtual environments: modern psychophysics in spatial cognition research. In: Jenkin M, Harris L (eds) Vision and attention. Springer, New York, pp 233–252

Cavanagh P, Labianca AT, Thornton IM (2001) Attention-based visual routines: sprites. Cognition 80:47–60

Chong SC, Treisman A (2005) Attentional spread in the statistical processing of visual displays. Percept Psychophys 67:1–13

Chun MM, Jiang Y (1998) Contextual cuing: implicit learning and memory of visual context guides spatial attention. Cogn Psychol 36:28–71

Chun MM, Jiang Y (1999) Top-down attentional guidance based on implicit learning of visual covariation. Psychol Sci 10:360–365

Chun MM, Jiang Y (2003) Implicit, long-term spatial contextual memory. J Exp Psychol Learn Mem Cogn 29:224–234

Clark RC (1999) Developing technical training. ISPI, Washington, DC

Cohen RA (1993) The neuropsychology of attention. Plenum Press, New York

Coull JT (1998) Neural correlates of attention and arousal: insights from electrophysiology, functional neuroimaging and psychopharmacology. Prog Neurobiol 55:343–361

Dalton P, Lavie N (2004) Auditory attentional capture: effects of singleton distractor sounds. J Exp Psychol Hum Percept Perform 30:180–193

Davenport J, Potter MC (2004) Scene consistency in object and background perception. Psychol Sci 15:559–564

Dennett D (2007) Heterophenomenology reconsidered. Phenomenol Cogn Sci 6:247–270

Di Lollo V, Smilek D, Kawahara J-I, Ghorashi SMS (2005) System reconfiguration, not resource depletion, determines efficiency of visual search. Percept Psychophys 67:1080–1087

Downing P, Liu J, Kanwisher N (2001) Testing cognitive models of visual attention with fMRI and MEG. Neuropsychologia 39:1329–1342

Embree L (ed) (2004). Preface. In: Gurwitsch's relevancy for cognitive science. Springer, Dordrecht

Engle RW, Tuholski SW, Laughlin J, Conway ARA (1999) Working memory, short-term memory, and general fluid intelligence: a latent variable approach. J Exp Psychol Gen 128:309–331

Eriksen CW, Schultz DW (1979) Information processing in visual search: a continuous flow conception and experimental results. Percept Psychophys 25:249–263

Eriksen CW, St. James JD (1986) Visual attention within and around the field of focal attention. Percept Psychophys 40:225–240

Fileteo JV, Freidrich FJ, Stricker JL (2001) Shifting attention to different levels within global-local stimuli: a study of normal participants and a patient with temporal-parietal lobe damage. Cogn Neuropsychol 18:227–261

Finkbeiner MA, Slotnick SD, Moo LR, Caramazza A (2007) Involuntary capture of attention produces domain-specific activation. NeuroReport 18:975–979

Folk CL, Remington RW, Johnston JC (1992) Involuntary covert orienting is contingent on attentional control settings. J Exp Psychol Hum Percept Perform 18:1030–1044

Folk CL, Remington RW (1998) Selectivity in distraction by irrelevant featural singletons. J Exp Psychol Hum Percept Perform 24:847–858

Frey H-P, König P, Einhäuser W (2007) The role of first- and second-order stimulus features for human overt attention. Percept Psychophys 69:153–161

Gibson BS, Kelsey EM (1998) Stimulus-driven attentional capture is contingent on attentional set for display-wide visual features. J Exp Psychol Hum Percept Perform 24:699–706

Glover S, Dixon P (2001) The role of vision in the on-line correction of illusion effects on action. Can J Exp Psychol 55:96–103

Gottesman CV, Intraub H (2002) Surface construal and the mental representation of scenes. J Exp Psychol: Hum Perform 28:1–11

Gurwitsch A (1964) The field of consciousness. Duquesne University Press, Pittsburgh, PA

Gurwitsch A (1966) Studies in phenomenology and psychology. Northwestern University Press, Evanston, IL

Gurwitsch A (1985) The phenomenology of signals and significations. In: Lester E (ed) Marginal consciousness. Ohio University Press, Athens, OH

Hanrahan C, Vergeer I (2001) Multiple uses of mental imagery by professional modern dancers. Imag Cogn Pers 20:231–255

Haywood J (1998) A rdzogs-chen Buddhist interpretation of the sense of self. J Conscious Stud 5:611–626

Henderson JM, Hollingworth A (1999) High-level scene perception. Annu Rev Psychol 50:243–271

Henderson JM, Hollingworth A (2003) Eye movements and visual memory: detecting changes to saccade targets in scenes. Percept Psychophys 65:58–71

Hershkowitz I, Orbach Y, Lamb ME, Sternberg KJ, Horowitz D (2001) The effect of mental context reinstatement on children's accounts of sexual abuse. Appl Cogn Psychol 15:235–248

Hicks JL, Marsh RL, Cook GI (2005) An observation on the role of context variability in free recall. J Exp Psychol Learn Mem Cogn 31:1160–1164

Holland RW, Hendriks M, Aarts H (2005) Smells like clean spirit: nonconscious effects of scent on cognition and behavior. Psychol Sci 16:689–693

Hollingworth A (2005) The relationship between online visual representation of a scene and long-term scene memory. J Exp Psychol Learn Mem Cogn 31:396–411

Hollingworth A, Hyun J-S, Zhang W (2005) The role of visual short-term memory in empty cell localization. Percept Psychophys 67:1332–1343

Hommel BK, Ridderinkhof R, Theeuwes J (2002) Cognitive control of attention and action: issues and trends. Psychol Res 66:215–219

Horowitz TS, Klieger SB, Fencsik DE, Yang KK, Alvarez GA, Wolfe JM (2007) Tracking unique objects. Percept Psychophys 69:172–184

Husserl E (1970) Logical investigations, vols. 1 and 2 (trans: John. N. Findlay). Routledge, London

Husserl E (1982) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. First book. General introduction to pure phenomenology (trans: Fred Kersten). Kluwer, Dordrecht

Husserl E (1991) On the phenomenology of the consciousness of internal time 1893–1917 (trans: John Barnett Brough). Kluwer, Dordrecht

Husserl E (2001) Analyses concerning passive and active synthesis: lectures on transcendental logic (trans: Anthony Steinbock). Kluwer, Dordrecht

Intraub H, Hoffman JE, Wetherhold JC, Stoehs S-A (2006) More than meets the eye: the effect of planned fixations on scene representation. Percept Psychophys 68:759–769

Irwin DE, Zelinsky GH (2002) Eye movements and scene perception: memory for things observed. Percept Psychophys 64:882–895

Jacobson NS, Dobson KS, Traux PA, Addis ME, Koerner K, Gollan JK, Gortner E, Prince SE (1996) A competent analysis of cognitive-behavioural treatment for depression. J Consult Clin Psychol 64:295–304

James W (1981) Principles of psychology, vol. 1. Harvard University Press, Cambridge, MA

Jescheniak JD, Hantsch A, Schriefers H (2005) Context effects on lexical choice and lexical activation. J Exp Psychol: Learn Mem Cogn 31:905–920

Jiang Y, Chun MM (2003) Contextual cuing: reciprocal influences between attention and implicit learning. In: Luis J (ed) Attention and implicit learning. John Benjamins, Amsterdam, pp 277–296

Jones MR, McAuley JD (2005) Time judgments in global temporal contexts. Percept Psychophys 67:398–417

Kambe G, Rayner K, Duffy SA (2001) Global context effects on processing lexically ambiguous worlds: evidence from eye fixations. Mem Cogn 29:363–372

Kanwisher N (2001) Neural events and perceptual awareness. Cognition 79:89–113

Kelso JAS, Case P, Holroyd T, Horvath E, Raczaszek J, Tuller B, Ding M (1995) Multistability and metastability in perceptual and brain dynamics. In: Kruse P, Stadler M (eds) Ambiguity in mind and nature. Springer-Verlag, Heidelberg

Kelter S, Kaup B, Claus B (2004) Representing a described sequence of events: a dynamic view of narrative comprehension. J Exp Psychol Learn Mem Cogn 30:451–464

Koch C, Crick F (1994) Some further ideas regarding the neuronal basis of awareness. In: Koch C, Davis JL (eds) Large scale neuronal theories of the brain. MIT Press, Cambridge, MA, pp 93–110

Koffka K (1925) Psychologie. In: Dessoir M (ed) Lehrbuch der Philosophie. Springer, Berlin

Kunzendorf R, Hartmann E, Thomas L, Berensen L (2000) Emotionally directing visual sensations: I. Generating images that contextualize emotion and become "symbolic." Imag Cogn Pers 19:269–278

LaBerge D (2002) Attentional control: brief and prolonged. Psychol Res 66:220–223

Lamy D, Segal H, Ruderman L (2006) Grouping does not require attention. Percept Psychophys 68:17–31

Lappin JS, Shelton AL, Rieser JJ (2006) Environmental context influences visually perceived distance. Percept Psychophys 68:571–581

Lewandowsky S, Kirsner K (2000) Knowledge partitioning: context-dependent use of expertise. Mem Cogn 28:295–305

Libera CD, Chelazzi L (2006) Visual selective attention and the effects of monetary rewards. Psychol Sci 17:222–227

Lippman LG, Sucharski IL, Bennington K (2001) Contextual connections to puns in fables: perceived humor. J Gen Psychol 128:157–169

MacGregor JN, Ormerod TC, Chronicle EP (2001) Information processing and insight: a process model of performance of the nine-dot and related problems. J Exp Psychol Learn Mem Cogn 27:176–201

MacLeod AK, Salaminiou E (2001) Reduced positive future-thinking in depression: cognitive and effective factors. Cogn Emotion 15:99–107

Malmberg KJ, Shiffrin RM (2005) The 'one-shot' hypothesis for context storage. J Exp Psychol Learn Mem Cogn 31:322–336

Mandel DR (2003) Counterfactuals, emotions, and context. Cogn Emotion 17:139–159

Marino AC, Scholl BJ (2005) The role of closure in defining 'objects' of object-based attention. Percept Psychophys 67:1140–1149

Marsh E, Edelman G, Bower GH (2001) Demonstrations of a generation effect in context memory. Mem Cogn 29:798–805

McAuley JD, Jones MR (2003) Modeling effects of rhythmic context on perceived duration: a comparison of interval and entrainment approaches to short interval timing. J Exp Psychol Hum Percept Perform 29:1102–1125

McNamara DS, McDaniel M (2004) Suppressing irrelevant information: knowledge activation or inhibition? J Exp Psychol Learn Mem Cogn 30:465–482

Mendolia M, Moore J, Tesser A (1996) Dispositional and situational determinants of repression. J Pers Soc Psychol 70:856–867

Merleau-Ponty M (1962) Phenomenology of perception (trans: Colin Smith). The Humanities Press, New Jersey

Metzinger T (2003) Phenomenal transparency and cognitive self-reference. Phenomenol Cogn Sci 2:353–393

Miller JO (1991) The Flanker compatibility effect as a function of visual angle, attentional focus, visual transients, and perceptual load: a search for boundary conditions. Percept Psychophys 49:270–288

Mischel W, Shoda Y, Rodriguez ML (1989) Delay of gratification in children. Science 244:933–938

Morgan JL, Meyer AS (2005) Processing of extrafoveal objects during multiple-object naming. J Exp Psychol Learn Mem Cogn 31:428-442

Mou W, Zhang K, McNamara TP (2004) Frames of reference in spatial memories acquired from language. J Exp Psychol Learn Mem Cogn 30:171–180

Navon D (1977) Forest before trees: the precedence of global features in visual perception. Cogn Psychol 9:441–474

Nelson DL, McEvoy CL, Pointer L (2003) Spreading activation or spooky action at a distance? J Exp Psychol Learn Mem Cogn 29:42–52

Newstead SE, Coventry KR (2000) The role of context and functionality in the interpretation of quantifiers. Eur J Cogn Psychol 12:243–259

Noles NS, Scholl BJ, Mitroff SR (2005) The persistence of object file representations. Percept Psychophys 67:324–334

Oberauer K (2002) Access to information in working memory: exploring the focus of attention. J Exp Psychol Learn Mem Cogn 28:411–421

Ohman A, Flykt A, Esteves F (2001) Emotion drives attention: detecting the snake in the grass. J Exp Psychol Gen 130:466–478

Olson I, Chun MM (2001) Temporal contextual cuing of visual attention. J Exp Psychol Learn Mem Cogn 27:1299–1313

Olson I, Marshuetz C (2005) Remembering "what" brings along "where" in visual working memory. Percept Psychophys 67:185–194

Palmer S (1999) Vision science: from photons to phenomenology. MIT Press, Cambridge, MA

Pasto L, Burack JA (2002) Visual filtering and focusing among persons with schizophrenia, major depressive disorder, and no psychiatric history. Can J Behav Sci 34:239–249

Peterson MA, Enns JT (2005) The edge complex: implicit memory for figure assignment in shape perception. Percept Psychophys 67:727–740

Posner M (1980) Orienting of attention. Q J Exp Psychol 32:3–25

Posner M (1995) Attention in cognitive neuroscience: an overview. In: Michael G (ed) The cognitive neurosciences. MIT Press, Cambridge, MA, pp 625–648

Pylyshyn ZW (2001) Visual indexes, preconceptual objects, and situated vision. Cognition 80:127–158

Pylyshyn ZW (2003) Seeing and visualizing: it's not what you think. MIT Press, Cambridge, MA

Quinn PC, Bhatt RS (2005) Good continuation affects discrimination of visual pattern information in young infants. Percept Psychophys 67:1171–1176

Rafal R, Henik A (1994) The neurobiology of inhibition: integrating controlled and automatic processes. In: Dagenbach D, Carr TH (eds) Inhibitory processes in attention, memory, and language. Academic, New York

Raidl M-H, Lubart TI (2001) An empirical study of intuition and creativity. Imag Cogn Pers 20:217–230

Rauschenberger R, Yantis S (2001) Attentional capture by globally defined objects. Percept Psychophys 63:1250–1261

Rawson K, Kintsch W (2002) How does background information improve memory for text content. Mem Cogn 30:768–778

Rees G, Lavie NN (2001) What can functional imaging reveal about the role of attention in visual awareness. Neuropsychologia 39:1343–1353

Rensink RA (2000) The dynamic representation of scenes. Visual Cogn 7:17–42

Rochat P (ed) (1999) Early social cognition: understanding others in the first months of life. Lawrence Erlbaum Associates, Mahwah, NJ

Roney CJR, Trick LM (2003) Grouping and gambling: a gestalt approach to understanding the gambler's fallacy. Can J Exp Psychol 57:69–75

Ruddle RA, Lessels S (2006) For efficient navigational search, humans require full physical movement, but not a rich visual scene. Psychol Sci 17:460–465

Russell PA (2000) The aesthetics of rectangle proportion: effects of judgment scale and context. Am J Psychol 113:27–42

Sartre J-P (1956) Being and nothingness (trans: Hazel Barnes). Philosophical Library, New York

Scholl BJ, Pylyshyn ZW (1999) Tracking multiple items through occlusions: clues to visual objecthood. Cogn Psychol 38:259–290

Scholl B, Pylyshyn ZW, Feldman J (2001) What is a visual object? Evidence from target merging in multiple object tracking. Cognition 80:159–177

Scholl B (2001) Objects and attention: the state of the art. Cognition 80:1–46

Sears CR, Pylyshyn Z (2000) Multiple object tracking and attentional processing. Can J Psychol 54:1–14

Sharps M, Martin SS (2002) "Mindless" decision making as a failure of contextual reasoning. J Psychol 136:272–282

Simons D, Mitroff SR (2001) The role of expectations in change detection and attentional capture. In: Michael J, Laurence H (eds) Vision and attention. Springer, New York, pp 189–207

Simons D, Wang RF, Roddenberry D (2002) Object recognition is mediated by extraretinal information. Percept Psychophys 64:521–530

Smith M, Benton S, Spalek T (2001) Attention constraints of semantic activation during visual word recognition. J Exp Psychol Learn Mem Cogn 27:1289–1298

Taube-Schiff M, Segalowitz N (2005) Linguistic attention control: attention shifting governed by grammaticized elements of language. J Exp Psychol Learn Mem Cogn 31:508–519

Terry WS, Burns JS (2001) Anxiety and repression in attention and retention. J Gen Psychol 128:422–432

Theeuwes J (1991) Cross-dimensional perceptual selectivity. Percept Psychophys 50:184–193

Theeuwes J (1994) Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. J Exp Psychol Hum Percept Perform 20:799–806

Theeuwes J (2004) Top-down search strategies cannot override attentional capture. Psychon Bull Rev 11:65–70

Theeuwes J, Chen CYD (2005) Attentional capture and inhibition (of return: the effect on perceptual sensitivity. Percept Psychophys 67:1305–1312

Tremblay S, Vachon F, Jones D (2005) Attentional and perceptual sources of the auditory attentional blink. Percept Psychophys 67:195–208

Troyer AK, Craik FIM (2000) The effect of divided attention on memory for items and their context. Can J Exp Psychol 54:161–170

Varela FJ, Thompson ET, Rosch E (1991) The embodied mind. MIT Press, Cambridge, MA

Varela F (1999) Present time consciousness. J Conscious Stud 6:405–428

Vecera SP, Flevaris AV, Filapek JC (2004) Exogenous spatial attention influences figure-ground assignment. Psychol Sci 15:20–26

Wallace BA (1999) The Buddhist tradition of Samatha: methods for refining and examining consciousness. J Conscious Stud 6:175–187

Waszak F, Drewing K, Mausfeld R (2005) Viewer-external frames of reference in the mental transformation of 3-D objects. Percept Psychophys 67:1269–1279

Watkins E, Teasdale JD, Williams RM (2003) Contextual questions prevent mood primes from maintaining experimentally induced dysphoria. Cogn Emotion 17:455–475

Watson DG, Humphreys GW (1997) Visual marking in moving displays: feature-based inhibition in not necessary. Percept Psychophys 63:74–84

Watson DG, Humphreys GW (2005) Visual marking: the effects of irrelevant changes on preview search. Percept Psychophys 67:418–434

Wertheimer M (1921) Untersuchungen zur Lehre von der Gestalt I. Psychologische Forschung 1:47–58

Williamon A, Valentine E, Valentine J (2002) Shifting the focus of attention between levels of musical structure. Eur J Cogn Psychol 14:493–520

Worthen JB, Wood VV (2001) A disruptive effect of bizarreness on memory for relational and contextual details of self-performed and other performed acts. Am J Psychol 114:535–546

Yang L-X, Lewandowsky S (2003) Context-gated knowledge partitioning in categorization. J Exp Psychol Learn Mem Cogn 29:663–679

Yantis S (1993) Stimulus-driven attentional capture. Curr Dir Psychol Sci 2:156–161

Zelinsky GJ, Loschky LC (2005) Eye movements serialize memory for objects in scenes. Percept Psychophys 67:676–690

Zwaan RA, Madden CJ (2004) Updating situations models. J Exp Psychol Learn Mem Cogn 30:283–288

Zwaan RA, Radvansky GA (1998) Situation models in language and memory. Psychol Bull 123:162–185

The Phenomenology and Neurobiology of Moods and Emotions

Matthew Ratcliffe

Introduction

Emotions are often conceived of by neurobiologists as states of bodily arousal that sometimes but not always lead to conscious feelings. It is assumed that, in those cases where the emotion is experienced, the experience takes the form of a *feeling of bodily changes*. In contrast, most philosophers regard emotions as conscious states which incorporate world-directed 'perception', 'construal' or 'judgement'. In this chapter, I first exemplify these opposing views by outlining the positions of Antonio Damasio, who takes emotions to be bodily changes that are sometimes felt, and Robert Solomon, who claims that they are judgements that are constitutive of world-experience. Then I raise some concerns about the practice of referring to non-conscious bodily states as 'emotions' and to experiences of these changes as 'feelings', after which I focus more specifically upon the view that *emotional experiences* are *feelings of certain kinds of bodily change*. My primary aim in what follows is to show that – if we set aside terminological differences – this view is not, after all, so different from Solomon's. In fact, the contrast between these two seemingly opposed positions is symptomatic of a mischaracterisation of the phenomenology of bodily feeling. This mischaracterisation can, I suggest, be corrected by drawing on the ideas of Husserl, Heidegger, Merleau-Ponty and other phenomenologists.

Damasio and Solomon on Emotion

Damasio's account of emotion is essentially an updated and elaborated version of something that was first suggested by William James (1884, 1890). According to James, emotions are feelings of physiological changes that are triggered reflexively during perception. Emotions, he suggests, regulate our behaviour by tuning us to the world in certain ways, steering us towards things that feel good and away from

M. Ratcliffe
Department of Philosophy, Durham University, UK

things that feel bad. Damasio similarly emphasises "bioregulatory reactions" that promote survival and well-being by guiding behaviour. These, he says, can be caused by stimuli in the external world or by conscious appraisals of situations. So a judgement about a situation is not part of an emotion but it can cause an emotion. According to Damasio, emotions consist of changes in both body and brain, although the bodily changes are sometimes replaced by simulated bodily changes, produced by what he calls an "as if" loop in the brain, which circumvents the rest of the body (1996, p. 1415). Central to his position is the "somatic marker hypothesis", the claim that the somatic (bodily) states which make up emotions are reliably associated with certain stimuli, "marking" those stimuli in such a way as to aid our decision-making. Somatic markers regulate decision-making and behaviour by making only certain parts of the environment perceptually salient and by classifying some of these as 'good' and others as 'bad' (Damasio 1995, p. 117). Emotions need not be consciously *felt* in order to play a role in reasoning. Non-conscious neurobiological processes can still regulate attention and contribute to what we sometimes call our *intuitions*. However, conscious emotional feelings do of course have a role to play too.

Damasio distinguishes between primary and secondary emotions. The former are innate sets of preferences or "preorganized mechanisms", whereas the latter are acquired associations between stimuli and states of bodily arousal, which utilise many of the same neural mechanisms as primary emotions (1995, p. 131). He adds that there is also an 'emotional background'. When no primary or secondary emotions are occurring, it is not that there is no emotion at all. Rather, there is an unobtrusive backdrop of bioregulatory processes (background emotion), the existence and consistency of which plays a vital role in shaping experience, thought and behaviour.

As well as offering a theory of the nature and role of emotions and feelings, Damasio offers a detailed account of the brain areas that are most likely involved. He claims that there is no single, dedicated 'emotion system' and that a number of areas are implicated in the causation of emotions:

> Sites such as the amygdala are part of multi-region systems that trigger emotions. On the other hand, sites in the hypothalamus, the basal forebrain (e.g., the nucleus accumbens), and in the brainstem (e.g., the nuclei in the perinqueductal gray), are the principal executors of an emotion. These are structures that directly signal, chemically and neurally, to the body and brain targets whose changes will come to constitute an emotional state. (2004, p. 54)

Damasio suggests that different areas are involved in the detection of emotion and in the generation of feeling, with the insula being "the key cortical component" implicated in emotional feeling (Damasio 2004, p. 55).[1]

Although James and Damasio agree that emotions involve bodily changes and that they regulate behaviour, their accounts differ in at least one important respect. For James, the emotion is the *feeling* of the relevant bodily changes. However, Damasio identifies the emotion with *non-conscious* changes in body and brain, changes that may or may not be felt. Feelings, he says, "*are the mental representation*

---

[1]See also Damasio (1996) for details of the relevant neurobiology.

of the physiologic changes that occur during an emotion" (2004, p. 53). It follows that phenomenological reflection has nothing whatsoever to tell us about emotion, given that emotions have no phenomenology. However, phenomenology can at least contribute to the description of feeling. Primary, secondary and background emotions are all sometimes felt. And Damasio indicates that it is background feelings which make the most important contribution to our experience:

> I am postulating another variety of feeling which I suspect preceded the others in evolution. I will call it *background feeling* because it originates in 'background' body states rather than in emotional states. It is not the Verdi of grand emotion, nor the Stravinsky of intellectualized emotion but rather a minimalist in tone and beat, the feeling of life itself, the sense of being. (1995, p. 150)[2]

Hence it is not only sudden eruptions of bodily arousal and strong feeling that drive us. There are also these less conspicuous feelings, which, he says, subtly "ebb and flow" but can also remain fairly constant for prolonged periods, thus contributing to a "mood". Damasio claims that a consistent background of bodily feeling gives us an anchor in the world, a point of relative stability in an ever-changing environment. It amounts, he suggests, to a core sense of self (1995, pp. 150–151). So, although he divorces emotion from phenomenology, he does reflect upon the phenomenology of feeling. Indeed, he indicates that feelings play a pivotal role in our lives, even constituting our 'sense of being'.[3]

Solomon offers a very different account of emotion. He stresses that emotions are not brief moments of bodily arousal or, for that matter, consistent backgrounds of arousal. Instead they are "intelligent, cultivated, conceptually rich engagements with the world" (1993, ix). It is the emotions that make our lives meaningful; "it is because we are moved, because we feel, that life has a meaning" (1993, ix).[4] Emotions, Solomon claims, are conscious judgements rather than non-conscious bioregulatory processes or conscious *feelings*. They are not judgements that *follow* our experiences and arise through a process of inference. Instead, they are *constitutive* of experience. The object of an emotion is experienced *through* the emotion and shaped by that emotion. Emotions therefore have a phenomenological structure; they are "*constitutive* judgments according to which our reality is given its shape and structure" (1993, xvii). They imbue the world with a significance that we ordinarily take for granted. Things appear to us as valuable, functional, important, worthwhile, pressing, and so

---

[2]See Damasio (2000, Chapter 9) for further discussion of background feeling.

[3]Although I will suggest that Damasio's conception of emotional feeling is plausible in at least some respects, I do not wish to endorse the way he uses the terms 'emotion' and 'feeling'. And what I will say here does not require me to accept the empirical details of his account, such as some of the specific claims he makes about the relationship between somatic markers and decision-making, and about the roles of particular brain areas. See Colombetti (2008) for an excellent critique of the 'somatic marker hypothesis'.

[4]I refer to the revised edition of Solomon's book *The Passions*, published in 1993. The book was first published in 1976 and played a key role in reinvigorating philosophical discussion of the emotions.

on. An experiential world that was not structured in such a way would be a place without significance, where nothing would show up as worthwhile, where nothing would draw us in or solicit us to act. So emotions are not objects of experience or ways in which only certain things are experienced. Rather, they comprise a background framework that shapes all experience; they "constitute our world, our relationships with other people and, consequently, our *Selves*" (Solomon 1993, p. 15).

Solomon claims that we are responsible for regulating our emotions just as we are responsible for regulating our beliefs, as emotions are complicated cognitive states that are amenable to rational critique. He also stresses that they are intricate processes which unfold over time, rather than brief neurobiological events. Against neurobiological approaches, he protests that "an emotion is *not* what happens in the first 120 ms of arousal" (2004b, p. 19). Solomon is also thoroughly dismissive of the view that emotions are non-conscious processes, as it runs contrary to everyday use of the term 'emotion' and therefore gives the misleading impression that a rich phenomenology, a significant world that we set up for ourselves and are responsible for rationally regulating, is nothing more than non-conscious arousal. He acknowledges the need to distinguish emotions from feelings but emphasises that this should not be done "in an arbitrary way that violates our most common understanding and renders the emotions beyond the reach of reflection and any semblance of (non-drug-induced) voluntary control" (Solomon 2007, p. 141).

In his earlier work, Solomon is equally dismissive of the view that emotions consist wholly or partly of bodily feelings. However, his more recent writings acknowledge that an adequate theory of emotion does need to accommodate the part played by the body in emotional experience. To achieve this, he appeals to the category of "kinaesthetic judgments" or judgements of the body (e.g. Solomon 2004b, p. 23). In our everyday practical activities, we make all sorts of non-conceptual 'judgements'. For example, in reaching out to catch a ball, the shape of one's hand and the way it moves comprise a judgement regarding the size, speed and weight of the ball. The bodily 'feelings' involved in emotions are, Solomon proposes, akin to such judgements. They are active engagements with the world, rather than passive states that are separate from judgement and action (see also Solomon 2003).

In distancing emotions altogether from passive bodily feelings, Solomon's position lies at one extreme end of a spectrum of views. The claim that emotions are neurobiological changes or feelings of such changes lies at the other.[5] It might be argued that concerns about referring to non-conscious biological processes as 'emotions' are merely terminological – different people are using the term 'emotion' in different ways and there is no substantive disagreement here. However, in some cases at least, there is more to it than that. Consider, for example, the approach laid out by LeDoux (1999). Like Damasio, he conceives of emotions as "biological functions of the nervous system" and argues that many of the behaviours associated with emotions are caused by

---

[5] Various philosophers have offered more conciliatory approaches, which acknowledge that emotions consist of both bodily feelings and world-directed, cognitive states. See the essays in Solomon ed. (2004a) for a range of philosophical positions.

these processes rather than by conscious feelings (p. 12). Given this, one might assume that the role of the phenomenologist is restricted to exploring the nature, role and variety of what LeDoux refers to as "conscious emotional feelings" (p. 17). However, he does not even allow this much. Instead, he claims that differences between kinds of emotion are differences between kinds of neurobiological process; the only difference between a conscious emotion and an unconscious emotion of the same type consists in the former's being conscious. And there is only the one kind of consciousness: "There is but one mechanism of consciousness and it can be occupied by mundane facts or highly charged emotions" (p. 19). Non-conscious emotions pass under the spotlight of consciousness and, when they do so, pre-existent differences between them are lit up. The differences that are lit up are the same differences that the neuro-biologist describes. So there is nothing for the phenomenologist to do when it comes to describing and differentiating kinds of emotional feeling.

I do not understand how anyone could be drawn to such a caricature of experience, where non-conscious biological processes *enter into* some amorphous realm called consciousness, which somehow reveals differences between them that are exactly the same differences as those studied by the neurobiologist. In fact, the only motivation I can think of for separating emotion and experience so decisively is that the neurobiologist can then claim sole rights to the former by dismissing altogether the possibility of any interesting questions arising with respect to specifically *emotional* experience. So Solomon is quite right to raise the concern that some approaches trivialise emotional experience and, in so doing, divert discussion of emotion away from a rich phenomenology and towards neurobiological arousal.

Damasio, however, is not so downbeat when it comes to phenomenology. Indeed, his 'background feelings' and Solomon's 'emotions' play very similar phenomenological roles. Both constitute the meaningfulness of a life, the sense of being a stable self that inhabits a significant world. Solomon (2003, p. 179) stresses that, unlike Damasio and LeDoux, he is "interested in the meanings of life, not short-term neurological arousal". But Damasio, at least, seems to be interested in both, and he and Solomon both stumble upon what looks like the same aspect of experience. In what follows, I will further clarify this aspect by turning to Heidegger's discussion of mood. Then I will draw upon the work of other phenomenologists in order to show how something can be both a bodily feeling and a sense of being situated in a world. Certain kinds of bodily feeling set up the meaningful world that we find ourselves in. So it turns out that Solomon and Damasio are gesturing at the same thing but using different terms.

Heidegger on Moods and Emotions

If experience and cognition are conceived of in terms of a detached subject impartially surveying the contents of an external world and contemplating their nature, then mood might look like a cognitive inconvenience. Mood gets in the way of cognition, by tainting experience and thought with a subjective colouration and thus impeding the ability to cognise the world in an objective fashion. However,

as Heidegger recognises, we are not voyeuristic subjects that gaze out upon a neutral, external realm. Rather, we find ourselves *in* the world; it is something that we implicitly take for granted as a context *within* which we experience, think and act. The experience of being part of a world is not a matter of being plonked into a space-time manifold and occupying a specific position in relation to a range of other objects. We are not *in* the world in the way that an object might be in a box. We are purposively entwined with the world, involved in it. It is a realm of practical significance and salient possibilities for activity that we ordinarily take for granted when pursuing our projects. Things call out for a range of activities and knit together with other things in intricate networks of teleological relations.

Heidegger claims that moods and other 'affective' states have been wrongly trivialised by philosophy since the time of Aristotle. They "sink to the level of accompanying phenomena" when in fact the affective aspects of experience, especially mood (*Stimmung*), play an indispensable role in structuring experience, thought sand activity (Heidegger 1962, p. 178). The world can only show up as it does in so far as things *matter* to us. And it is moods that constitute a sense of things mattering. Mood, for Heidegger, is not a kind of psychological state that we experience within an already given world. Mood is a background through which it is possible to encounter things in the ways that we do, as 'there', 'mattering', 'not mattering', 'for this' or 'for that'. Heidegger uses the term *Befindlichkeit* to refer to the way in which moods comprise *ways of finding oneself in a world*. It is translated in a number of ways but I favour 'attunement', a term suggested by Stambaugh in her translation of Heidegger's *Being and Time* Heidegger (1996). Moods, for Heidegger, attune us to the world; they give us a sense of being there, of being amongst things, harmoniously connected to things.

Heideggerian moods have neither an 'internal' nor an 'external' phenomenology. Any contrast between internal or subjective mental states and an external or objective world presupposes the background sense of belonging to a world that is constituted by mood. We find ourselves *in* a mood, rather than experiencing ourselves as subjects who *have* moods: "A mood assails us. It comes neither from 'outside' nor from 'inside', but arises out of Being-in-the-world, as a way of such Being" (Heidegger 1962, p. 176). So a mood is not a kind of intentional state, directed at the body or at anything else. And neither is it some kind of non-intentional bodily feeling. A mood is a background to all specifically directed intentional states. It is part of the structure of intentionality and is presupposed by the possibility of encountering anything in experience or thought: "*The mood has already disclosed, in every case, Being-in-the-world as a whole and makes it possible first of all to direct oneself towards something*" (1962, p. 176). We are never free of moods – a mood is not an occasional eruption. Moods do of course change but, when they do, one mood is replaced by another mood rather than by no mood at all (Heidegger 1962, p. 175). The everyday 'lack of mood', the experience of going about one's business in a way that is uninterrupted by a 'good' or 'bad' mood is itself a mood, a backdrop through which the world shows up in the usual, familiar, unremarkable way.

Heidegger does not discuss, in any detail, the range of moods we experience. In addition to the nondescript, everyday mode of dwelling in the world, he mentions some varieties of fear, including dread, for example, which he takes to be a

fear of the unfamiliar (1962, p. 182). He also mentions elation (1962, p. 182) and, in some slightly later works, he discusses the phenomenology of boredom (1978, Heidegger 1983). It might seem peculiar that he refers to fear as a mood. Surely fear is not a mode of attunement but a specifically focussed emotion, an *experience of something in a world* rather than a *way of being in the world*? However, the 'mood' of fear to which Heidegger refers is not a specifically directed intentional state, such as 'fear of the raging bull'. He is referring to the mood that is itself the *possibility of fear*. In order to be afraid of something, one must already find oneself in the world in a way that incorporates certain kinds of self-concern. A being that was indifferent to its own being could not be afraid:

> Fearing, as a slumbering possibility of Being-in-the-world in [an attunement] […] has already disclosed the world, in that out of it something like the fearsome may come close. […] Only an entity for which in its Being this very Being is an issue, can be afraid. Fearing discloses this entity as endangered and abandoned to itself. (1962, p. 180)

Moods are not generalised emotions but conditions of possibility for specifically focused emotions such as *fear of something*. The range of occurrent emotions that can be experienced is determined by the shape of the background mood. A being without any self-concern, a being that did not *care* for its being in some pre-conceptual, felt way, would not be open to the possibility of emotions such as fear.

A mood that Heidegger does discuss at length is *Angst* or anxiety. Not all moods constitute a sense of being *at home* in the world, of being in a familiar, significant realm where things matter. Sometimes we experience a kind of global experiential defamiliarisation, where everything appears *unheimlich* or 'uncanny'. The background of ordinarily taken-for-granted familiarity ebbs away and is replaced by an all-enveloping sense of unfamiliarity and disconnectedness. Angst is the most extreme version of this, a complete absence of practical significance. Heidegger refers to this as an experience of the 'nothing', as it is an experience of the total loss of world-meaning. The possibility of things being encountered as 'there', as part of a world in which one dwells, is gone. And this is because the world in which things show up, the usual backdrop of significance, is itself gone:

> The receding of beings as a whole that closes in on us in anxiety oppresses us. We can get no hold on things. In the slipping away of beings only this 'no hold on things' comes over us and remains. Anxiety reveals the nothing. (Heidegger 1978, p. 101)

All that remains is the sense of loss, of the entire structure of being-in-the-world slipping away. It follows that moods are not only responsible for a sense of *belonging* to the world but also for a sense of the *being* of things.

The Phenomenology of Feeling

Heidegger's discussion of mood complements some of Damasio's claims regarding the role of background feeling. In both cases, we have something that constitutes the sense of being in a meaningful world, the diminishment or alteration of which

affects us in profound ways. Needless to say, the two kinds of enquiry are very different. Heidegger is concerned with how it is that a meaningful experiential world is intelligible to us. Damasio, in contrast, is interested in neurobiological, *causal* processes and how they regulate our experience, deliberation and behaviour. The question of how a neurobiological account of background feeling might relate to a phenomenological description of the same thing is not an easy one to answer. Scientific enquiry presupposes a sense of there being an actual world, the contents of which can be encountered, systematically explored and described. Hence, when it comes to mood, science is exploring the neurobiological correlates of something that science at the same time takes for granted. It is not at all clear whether one approach has priority over the other or how the two can be united into a single, coherent conception of our relationship with the world. Nevertheless, they do at least both start off by acknowledging much the same phenomenology, a pre-conceptual, felt sense of being and belonging. And, in doing so, they at least point to the potential for fruitful dialogue between phenomenological and neurobiological enquiry.

There is a much closer connection between Heidegger's account of mood and Solomon's view, given that Solomon (1993) explicitly acknowledges Heidegger as a strong influence upon his thinking. And it is helpful to return to Heidegger in order to clarify a phenomenological distinction that Solomon does not make sufficiently clear. The experiential role that Solomon assigns to 'emotion' is essentially that of Heideggerian 'mood' – emotions set up the world we reside in. However, Solomon's discussion is sometimes confusing, as it neglects to draw a clear distinction between the background web of 'emotion' that opens up a meaningful world and those emotions that take the form of occurrent judgements. Some emotions happen within the already experienced world, whereas others comprise a sense of being-in-the-world that is presupposed by within-world emotions. Solomon regards moods as "generalized emotions" (1993, p. 15). But a generalised emotion no more constitutes a sense of belonging to the world than does fear of a fast-approaching car. Both are directed at things in the world, even though one is directed at a much wider range of things. However broadly directed an emotion might be, it still presupposes a sense of being *there*, of the possibilities that must already be in place for an object-directed emotion or more broadly directed mood to be possible. This is something that Heidegger succeeds in conveying. World-constituting moods are neither specifically focussed emotions nor broadly focussed emotions. Rather, they comprise a space of experiential possibilities that these and all other intentional states take for granted, the meaningfulness of life.

Conversely, I think that Solomon succeeds in drawing attention to a weakness in Heidegger's view. In *Being and Time*, much of Heidegger's discussion is preoccupied with the contrast between two opposing existential predicaments, the everyday background of mood and an anxiety that reveals the 'nothing'. This might give the impression of there being a linear continuum between everyday belonging and its total absence. But this construal of experience is misleading. Solomon discusses references to the 'Absurd' in Camus and others, and suggests that it is not – as is often claimed – an absence of meaning but instead a permutation of meaning.

As he puts it, it is "not meaninglessness at all, but a certain kind of meaning" which is symptomatic of the "self-demeaning view of ourselves" and found in predicaments such as depression (1993, pp. 50–51).

Much the same can be said of Heidegger's 'nothing'; it is not at all clear that it is the experience of an all-enveloping *loss* of world-meaning. Our moods or background feelings do not simply constitute *either* a background of oblivious belonging *or* an estrangement somewhere on the road to full-blown nothingness. In everyday life, people attempt to convey a wide range of subtly different existential predicaments, which differ not just in intensity but also in character. People do not simply belong to the world or feel estranged from it; they belong to it in different ways and they feel estranged in different ways too. For example, someone might talk of feeling alive, dead, distant, detached, dislodged, estranged, isolated, otherworldly, indifferent to everything, overwhelmed, suffocated, cut off, lost, disconnected, out of sorts, not oneself, out of touch with things, out of it, not quite with it, separate, in harmony with things, at peace with things or part of things. There are references to feelings of unreality, heightened existence, surreality, familiarity, unfamiliarity, strangeness, isolation, emptiness, belonging, being at home in the world, being at one with things, significance, insignificance, and the list goes on. Some of these might well be synonyms for others. Even so, the many different ways in which people describe how they find themselves in the world is suggestive of a variety of subtly different existential predicaments.

I refer to these as 'existential feelings' rather than as 'moods' (e.g. Ratcliffe 2005b, 2008a). They are 'existential', as they amount to different ways of being in the world. And I refer to them as 'feelings' rather than 'moods' for three reasons. First of all, 'feeling' is the term that people most often use when communicating such predicaments. Hence focusing on 'mood' tends to direct enquiry towards an overly restricted list of existential predicaments. Most of the existential feelings that people attempt to convey are not to be found on standard inventories of moods or emotions, and many do not even have established names. Second, many of them are brief episodes or dynamic processes, rather than enduring 'moods'. The third – and most important reason – for using the term 'feeling' is they are indeed 'bodily feelings'. Something can be both an existential orientation, what Solomon might call a set of world-constituting judgements, and at the same time a way in which the body feels. We can draw upon the work of phenomenologists to make this clear and, in so doing, dispense with the contrast between world-directed intentional states and internally-directed bodily feeling, a contrast that has been central to most recent discussion of emotions, feelings and moods.

We can draw upon a wide range of phenomenological descriptions to illustrate (a) the intimate connection between changed feeling and a changed sense of belonging to the world and (b) the range of different ways in which people can find themselves in the world. Some of the most vivid examples relate to altered experience in psychiatric illness. For example, William James offers the following description of what is now known as 'depression':

> In certain forms of melancholic perversion of the sensibilities and reactive powers, nothing touches us intimately, rouses us, or wakens natural feeling. The consequence is the complaint so often heard from melancholic patients, that nothing is believed in by them as it used to be, and that all sense of reality is fled from life. They are sheathed in india-rubber; nothing penetrates to the quick or draws blood, as it were. (1890, p. 298)

This suggests a changed way of belonging to the world, a loss of significance and an erosion of the usual sense of the reality of things, all of which are bound up with altered feelings of the body. The body that is sheathed in india-rubber is the same body that no longer finds itself *there*, in contact with things.[6]

Another example is the phenomenon of depersonalisation. This is a fairly common complaint, which is reported by around seventy percent of psychiatric in-patients and is sometimes but not always associated with other psychiatric conditions. It is also something that most of us experience, to a lesser degree, from time to time – in fatigue, jet lag, illness or maybe during a very bad hangover. Descriptions of the experience alternate between references to self, body, world, and the relationship between them. And there is a consistent emphasis on changed feeling; a loss of certain feelings is at the same time a loss of connectedness to the world and of the usual sense of reality. Medford et al. (2005, p. 93) describe some of the symptoms as follows:

> [S]ome patients report feeling 'like a robot, 'different from everyone else' and 'separate from myself' […] Others describe feeling 'half-asleep' or 'as if my head is full of cotton wool' […] External reality may also be strangely affected: it may appear somehow artificial – as if 'painted, not natural', or 'two-dimensional' or 'as if everyone is acting out a role on stage, and I'm just a spectator'. […] A reduction in, or complete absence of, bodily feelings is often described ('as if I were a phantom body', 'my hands seem not to belong to me'). Another frequent theme is a reduction or loss of emotional responses: 'my emotions are gone, nothing affects me', 'I am unable to have any emotions. Everything is detached from me'.[7]

The inextricability of feeling and world-experience is not adequately acknowledged by philosophical approaches that impose, from the outset, a crisp distinction between bodily feeling and world-directed intentionality. Most philosophers admit that emotions incorporate both world-directedness and bodily feeling but they construe the two as separate ingredients (e.g. Lyons 1980). Some have argued that feelings can be world-directed. But, in so doing, they still retain the internal–external contrast and so fail, to some degree at least, to respect the relevant phenomenology. For example, Prinz (2004) argues that feelings can be *about* things other than the body but he adopts a non-phenomenological conception of intentionality

---

[6] Contrary to orthodox interpretations of James's position, he does not claim that emotions are feelings of bodily changes, which are distinct from world-directed intentionality, but that they are bodily changes which are at the same time ways in which the world is experienced (see Ratcliffe 2005a). Indeed, in his later work, James explicitly states that the phenomenology of feeling is neither 'external' nor 'internal' in character and that it does not conform to such interpretive categories (see, e.g., James 1905).

[7] A variety of existential changes are also implicated in schizophrenia, many of which are addressed by Sass (e.g., 2004). See also Ratcliffe (2008a, Chapter 7).

and continues to assume that the phenomenology of feeling is internal in character. Goldie (2000), in contrast, does engage with the phenomenology. He claims that, in addition to having internally directed feelings, we also have 'feelings towards' – feelings that are at the same time world-directed intentional states. However, he distinguishes 'bodily feelings' from 'feelings towards', thus retaining a phenomenological distinction between 'internally' and 'externally' directed feelings.

Damasio too takes the phenomenology of feeling to be a phenomenology of the body. For example, he refers to background feeling as "our image of the body landscape when it is not shaken by emotion" (1995, pp. 150–151). If the feeling body is conceived of in this way, as an object of experience, it is quite unclear how these feelings can add up to a sense of 'being'. The felt body is not experienced as the source of all world-meaning. When we encounter our bodies as objects of experience, we are already in a world. We take for granted a context in which things, including our bodies, can be encountered. What is needed, therefore, is a different conception of the phenomenology of feeling, which acknowledges that it need not involve the body appearing as an *object* of experience. What Damasio fails to recognise or at least to make explicit is the possibility of bodily feeling being a *medium* through which something else is experienced, rather being itself an object of feeling (Gallagher 2005, pp. 135–137; Sass 2004, p. 134). The *feeling* body is implicated in all experience but not always as the *felt* body.

Heidegger (1962) claims that moods are neither internal nor external in character; they do not respect such distinctions. In so far as moods are feeling states, he thus challenges the assumption that they have an internal, bodily or purely subjective phenomenology.[8] However, it is not clear whether or not Heidegger does think of moods as having an essentially bodily phenomenology, given that the phenomenology of the body is not addressed in *Being and Time*. However, he does discuss the body in his *Zollikon Seminars*, which were held at the home of the psychiatrist Medard Boss between 1959 and 1969, and translated into English in 2001. Here, Heidegger quotes himself in *Being and Time* as acknowledging that the body "hides a whole problematic of its own", and stresses that the body, although neglected by his previous works, is important and needs to be discussed by the phenomenologist (Heidegger 2001, p. 80). He goes on to remark that in the course of our everyday activities the body is not a conspicuous object of experience and instead disappears into the background. Hence a phenomenology of the body will be, in part, a phenomenology of this disappearance, this peculiar absence. He also reflects upon the experience of touch.[9]

---

[8] See also Strasser (1977, Chapter 7) who, like Heidegger, addresses background mood (*Stimmung*). Strasser claims that moods do have a bodily phenomenology; they are a matter of feeling. He stresses that the relevant phenomenology cannot be captured in terms of a distinction between 'internal' and 'external', and claims that feeling is an opening onto the world, a way of belonging to it.

[9] The 'Conversations with Medard Boss', which are also included in the 2001 text, include further references to the body. For example, Heidegger states that the body is not "present-at-hand" (p. 170). In other words, experience of it is quite unlike spectatorial contemplation of an object. There are also some obscure references to "bodying forth" (p. 200), which leave Medard Boss somewhat bemused.

The disappearance of the body as an *object* of experience and also the phenomenology of touch are described more clearly and in more depth by both Husserl (e.g. Husserl 1989) and Merleau-Ponty (e.g. Merleau-Ponty 1962). For Husserl and Merleau-Ponty, the body is that through which we experience the world, rather than just an object of perception: "The Body (*Leib*) is, in the first place, the *medium of all perception;* it is the *organ of perception* and is *necessarily* involved in all perception" (Husserl 1989, p. 61).

But how can a bodily feeling be – at the same time – a way in which something else is experienced? We can show that this is at least sometimes the case by appealing to the phenomenology of touch (Ratcliffe 2008b). Much of current Anglophone philosophy of mind presupposes a somewhat spectatorial conception of experience, modelled upon a questionable conception of the visual modality. We 'look out' upon an external world and perceive its contents in a way that is seemingly distinct from how our bodies might be feeling. Hence 'internal' feelings get separated from world-experience. The phenomenology of touch is quite different from this. If we turn to touch, it is far from clear that bodily experience can be separated from world experience, as both Husserl (1989, Section II) and Merleau-Ponty (1962, Part II, Chapter 3) recognise. When you touch something very hot, the primary object of experience might well be your hand, as you pull it back in pain. However, in routine activities, where things proceed in accordance with our expectations, what is felt is not the hand but what it touches. The touch is a medium through which something else is perceived. The body continues to feel but is not itself a conspicuous object of feeling.

Of course, it could be argued that there is no bodily feeling at all in cases where something other than the body is perceived through touch; there is just the externally directed feeling. However, this is implausible. Consider a hand that has gone numb, after one has slept upon one's arm. One is aware of the absence of feeling and so the hand becomes a phenomenologically conspicuous *object* of experience. The returning feeling takes the form of the disappearance of the hand, which again becomes a medium of perception. It is felt again but not in an object-like way. One might retort that the hand does not *feel* again. Rather, it has disappeared from experience. But suppose you could implant a switch in your wrist, which – when activated – blocked all neural traffic without causing any pain or discomfort. Now suppose that you gave this switch to someone else, who turned it on and off without your knowing when. The question is whether you would feel any difference between the hand as it is usually felt (or not felt) and the hand from which all neural feedback was absent. And the answer is, I suggest, that you would, even if you were not doing anything with the hand, which was just hanging limply by your side. When a hand is not an object of perception but is instead an organ of perception or potential perception that hides in the background of experience, there remains a phenomenological difference between how it is experienced and not feeling anything at all.

It is also important to distinguish *localised* disappearance of parts of the body during tactile exploration from the tactile *background*. Specific parts of the body fall into the background as they become effortlessly immersed in activity. But,

when this occurs, much of the rest of the body surface also fades into the tactile background. The body as a whole is not something that we experience as set apart from the world. Background touch involves a lack of differentiation between body and world – bodily comfort is at the same time a failure to distinguish one's body from the world (Ihde 1983). So neither active nor passive touching usually draws clear experiential boundaries between body and world or between internal and external. The feeling is neither external nor internal in its phenomenology – it is a matter of connectedness, relatedness, lack of differentiation, which should not be pulled apart and re-interpreted in terms of two separate phenomenological components.

Hence, by reflecting upon the phenomenology of touch, we can progress at least some way to the view that bodily feelings are not phenomenologically distinct from a sense of belonging to the world or from specifically directed experiences of things in the world. However, in order to convey the extent to which a wide range of bodily dispositions – which are in some sense *felt* – permeate world-experience, I will turn to another theme in the work of Husserl and Merleau-Ponty, that of the *horizonal* structure of experience.

Horizons and Bodily Dispositions

As mentioned earlier, Solomon attempts to reconcile judgements with emotional feelings by incorporating the latter into the category of 'bodily judgements'. This might seem implausible, given that many of the visceral and other feelings involved in emotion are quite unlike the bodily dispositions and dynamics involved in catching a ball or judging the height of a stair. However, I suggest that a more detailed description of the role played by bodily dispositions in structuring experience, offered by Husserl (e.g. Husserl 1989, 2001) and Merleau-Ponty (e.g. Merleau-Ponty 1962), is able to accommodate the world-directedness of such feelings.[10] I will focus specifically upon the discussion in Husserl (2001).[11]

Husserl starts by noting that, when we perceive an object visually, we experience it *as* an enduring object residing in a public world. The object is "naturally and simply there for us as an existing reality as we live naively in perception" (2001, p. 35). Yet it is unclear how this is possible, given that all we see of the object is one aspect of it from one vantage point. Nevertheless, we do indeed see the *object*, rather than seeing a two-dimensional, synchronic snapshot that is then inferred to be a view of an object. So how can we experience the whole thing when we see only part of it?

---

[10] See also Husserl (1989, Section I, Chapter 2).

[11] I certainly do not want to suggest that Merleau-Ponty's view is exactly the same as Husserl's. However, most of the broad outline I sketch here is endorsed by both.

The answer is pretty simple – our experience is not restricted to the *actual*; we live in a space of *possibilities* and everything we experience is shaped by possibilities. For example, when you see a cup, your sense of it as a cup and as a real cup that is there right now does not consist only in the aspect that is currently presented. You also experience a range of possibilities involving the cup. For example, there are the possibilities of gazing upon it from a different angle so as to reveal its hidden aspects and of picking it up. And the possibility of drinking from it is especially salient. So, when we look at something, we also experience other potential ways in which it might be encountered and, in conjunction with these, ways in which it would appear:

> Noetically speaking, perception is a mixture of an actual exhibiting that presents us in an intuitive manner what is originally exhibited, and of an empty indicating that refers to possible new perceptions. In a noematic regard, what is perceived is given in adumbrations in such a way that the particular givenness refers to something else that is not-given, as what is not given belonging to the same object. (Husserl 2001, p. 41)

The possibilities that contribute to a sense of the existence and nature of a thing are not a disparate collection of isolated individuals, brought together arbitrarily in perception. Possibilities are interconnected in intricate, structured ways. They form "entire indicative systems" (Husserl 2001, p. 42) and the term that Husserl and Merleau-Ponty use to refer to such a system is "horizon".[12] Horizons are not static auras that surround frozen, experiential snapshots. Experience has a temporal structure and it is the harmoniously changing framework of possibilities that constitutes our sense of things in the world being enduring entities of various kinds.

Husserl and Merleau-Ponty also stress that experience is inter-modal, rather than being a matter of what is revealed to only one sense. What I see does not just offer up possibilities of seeing it from other vantage points. It may also incorporate salient possibilities for touch, smell, taste, audibility or practical manipulation. Actualities for one sense are possibilities for others and the possibilities for different senses combine to comprise the experience of something as *there*, as part of the world.

> … as the process of external perceiving progresses optically, the thing is not only intended optically. Intentions of the other sense spheres are continually co-awakened and must continually accord in the unity of a synthesis with the genuine impressional ones of the optical sphere. They must do this because they are co-constitutive of the objective sense. (Husserl 2001, p. 144)[13]

What does all this have to do with bodily feeling? In any experience, only certain possibilities are offered up and only some of these appear especially salient. According to both Husserl and Merleau-Ponty, the possibilities that show up are constituted by bodily dispositions. These dispositions shape all experience and show up as potentialities that belong to objects. The different ways in which the body

---

[12] Recent work on enactive perception (e.g. O'Regan and Noë 2001) complements this view in many respects. See Ratcliffe (2008a, Chapter 4) for a discussion.

[13] The sense of something as 'there', as 'real', also incorporates an appreciation of the possibilities that it offers for other people, of its not being available only to oneself (see Ratcliffe 2008a, Chapters 4–7).

responds to things amount to different systems of concrete possibilities. The active body thus operates as a medium of perception that is implicated in all experience; "the lived-body is constantly there, functioning as an organ of perception" and it features in our phenomenology primarily as an "I can" (Husserl 2001, pp. 50–51).

Possibilities, Husserl recognises, can have different degrees and kinds of appeal. Some are "open"; they are there but they do not draw us in. Others however are "enticing"; they have an "affective force" (p. 83). Pronounced bodily dispositions involving the object appear in the object and we feel the "affective pull of enticing possibilities" (p. 98). So Husserl is suggesting that feelings of the body manifest themselves as salient possibilities offered up by things in the world.

I do not think there any good grounds for restricting the relevant kinds of bodily feeling so as to exclude those that are characteristic of emotional states. Feelings in general are bound up with world-directed bodily dispositions. Even a stomach ache affects the possibility space – as the body becomes more conspicuous and pained, it becomes more salient as an object of perception and ceases to direct itself towards other things in the way it previously did. Hence there is a change in the possibilities that show up. Different kinds of feeling could be tied up with a range of bodily dispositions, such as fleeing, guarding oneself, hiding, approaching, exploring. And all of these are integral to the way in which things other than the body are experienced. Not every aspect of the horizon-structure of experience is a matter of specifically 'emotional' or 'affective' feeling. Nevertheless, these feelings do at least make some contribution, and a significant one at that.

The claim that bodily feelings structure experiences in this fashion applies equally to specifically directed feelings and to background 'existential' feelings. Husserl and Merleau-Ponty distinguish between the horizons that surround experienced objects and the more general world-horizon or universal horizon that such experiences presuppose. Suppose that a particular object appears tangible. It can only appear tangible, or intangible for that matter, in so far as tangibility is itself a possibility. The same applies to other kinds of possibility, such as that of being looked upon or grasped by another person, being hurt, being seen from another vantage point, and so on. All experiences occur against the backdrop of a general space of possibilities, a space of ways in which things can appear. This universal horizon – a possibility space that is also a set of bodily dispositions – is comparable to the background orientation that Heidegger says is constituted by mood. As Husserl describes it:

> All particular syntheses, through which things in perception in memory, etc., are given, are surrounded by a general milieu of empty intentions being ever newly awakened; and they do not float there in an isolated manner, but rather, are themselves synthetically intertwined with one another. For us the universal synthesis of harmonizing intentional syntheses corresponds to 'the world' and belonging to it is a universal belief-certainty. (2001, p. 146)

And, as Merleau-Ponty similarly puts it:

> The natural world is the horizon of all horizons, the style of all possible styles, which guarantees for my experiences a given, not a willed, unity underlying all the disruptions of my personal and historical life. Its counterpart within me is the given, general and prepersonal existence of my sensory functions … (1962, p. 330).

Certain kinds of bodily disposition are felt *as* how one finds oneself in a world. Background feelings and world-constituting judgements turn out to be one and the same. Of course, the question of whether we are responsible for the ways in which we find ourselves in the world, as suggested by Solomon, is another matter. It is, I think, unlikely that background feeling or world-constituting judgement is amenable to rational intervention to quite the extent that Solomon suggests or in so direct a way. Nevertheless, there are many ways in which we might act so as to regulate and change this background, such as behaving in a particular fashion, changing what we perceive, medicating ourselves or being selective about which people we talk to. Some such interventions can be carried out in a skilful fashion, making us far from passive before the world that we are thrown into.

Conclusion

I have suggested that, by drawing on the work of phenomenologists, we arrive at the view that certain bodily feelings are at the same time our sense of belonging to a world. These feelings need to be distinguished from other, more focused feelings that arise within an already experienced world. Together, these feelings shape our experiences of the world and things in it by revealing possibilities. When it comes to both world-constituting 'moods' or 'existential feelings' and to specifically focused emotions, the imposition of a division between internally directed feelings and externally directed experience is misleading, serving to split a single experience into two separate components.

However, I do not want to suggest that all specifically focused emotions are simply 'feelings'. Many 'emotions' are most likely complicated states, which incorporate a wide range of feelings, in addition to conceptualised appraisals and intricate narratives. In addition, emotions are often multi-faceted processes that unfold over lengthy periods of time (see, e.g. Goldie 2000). Neither do I discount the possibility that the term 'emotion' refers to a heterogeneous group of phenomena, some of which may be feelings, while others incorporate feelings but also have additional features (e.g. Griffiths 1997). But what I do want to maintain is that, in so far as emotions do involve feelings, those feelings should not be construed as distinct from the intentionality of emotion, as having an exclusively bodily phenomenology. Hence approaches which discount the possibility of emotions being bodily feelings *on the basis that* emotions are about things in the world whereas feelings are perceptions of bodily states are misconceived.

A question still to be answered is that of how a phenomenological approach might be reconciled with an account of the relevant neurobiology. One view is that the phenomenology is ultimately to be wholly explained in terms of a complete account of the relevant neurobiological mechanisms. However, I discount this on the basis that what Heidegger calls mood is presupposed by the *intelligibility* of the world that science takes for granted. As Heidegger remarks in the *Zollikon Seminars*, "one must see that science as such (i.e., all theoretical-scientific knowledge)

is founded as a way of being-in-the-world – founded in the bodily having of a world" (p. 94).

The world we live in is a world of the possible. The universal horizon, the space of possibilities that is presupposed by all encounters with things in the world, cannot be accommodated by an account that restricts itself to the actual. But this does not rule out mutually informative dialogue between phenomenology and science on the topic of moods, feelings and emotions, so long as that dialogue is not shaped from the outset by the assumption that we should be aspiring towards a state of affairs where the science ultimately trumps the phenomenology. In so far as the relevant phenomenology is presupposed by the possibility of aspiring towards an account of how the actual world is, in so far as it discloses the world that science then seeks to describe, it can never be wholly understood in terms of such a science.

References

Colombetti G (2008) The somatic marker hypotheses: what the Iowa Gambling Task does and does not show. British Journal for the Philosophy of Science, 59:51–71

Damasio A (1995) Descartes' Error: Emotion, Reason and the Human Brain. London: Picador

Damasio A (1996) The somatic marker hypothesis and the possible functions of the prefrontal cortex. Philos Trans R Soc Lond, Series B (Biol Sci) 351:1413–1420

Damasio A (2000) The feeling of what happens: body, emotion and the making of consciousness. Vintage, London

Damasio A (2004) Emotions and feelings: a neurobiological perspective. In: Manstead SR, Frijda N, Fischer A (eds) Feelings and emotions: the Amsterdam symposium. Cambridge University Press, Cambridge, pp 49–57

Gallagher S (2005) How the body shapes the mind. Oxford University Press, Oxford

Goldie P (2000) The emotions: a philosophical exploration. Clarendon, Oxford

Griffiths P (1997) What emotions really are: the problem of psychological categories. Chicago University Press, Chicago, IL

Heidegger M (1962) Being and time (trans: Macquarrie J, Robinson E). Blackwell, Oxford

Heidegger M (1978) What is metaphysics? (trans: Krell DF). In: Heidegger M, Krell DF (eds) Basic writings. Routledge, London, pp 93–110

Heidegger M (1983) The fundamental concepts of metaphysics. Indiana University Press, Bloomington, IN

Heidegger M (1996) Being and time (trans: Stambaugh J). State University of New York Press, New York

Heidegger M (2001) Zollikon Seminars: Protocols – Conversations – Letters (ed. Boss M; trans: Mayr F, Askay R). Northwestern University Press, Evanston

Husserl E (1989) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: second book (trans: Rojcewicz R, Schuwer A). Kluwer, Dordrecht

Husserl E (2001) Analyses concerning passive and active synthesis: lectures on transcendental logic (trans: Steinbock AJ). Kluwer, Dordrecht

Ihde D (1983) Sense and significance. Humanities, Atlantic Highlands, NJ

James W (1884) What is an emotion? Mind 9:188–205

James W (1890) The principles of psychology. Volume ii. Holt, New York

James W (1905) The place of affectional facts in a world of pure experience. J Philos Psychol Sci Methods II/11:281–287

LeDoux J (1999) The emotional brain: the mysterious underpinnings of emotional life. Phoenix, London

Lyons W (1980) Emotion. Cambridge University Press, Cambridge

Medford N, Sierra M, Baker D, David AS (2005) Understanding and treating depersonalisation disorder. Adv Psychiatr Treat 11:92–100

Merleau-Ponty M (1962) Phenomenology of perception (trans: Smith C). Routledge, London

O'Regan JK, Noë A (2001) A sensorimotor account of vision and visual consciousness. Behav Brain Sci 24:939–1031

Prinz J (2004) Gut reactions: a perceptual theory of emotion. Oxford University Press, Oxford

Ratcliffe M (2005a) William James on emotion and intentionality. Int J Philos Stud 13:179–202

Ratcliffe M (2005b) The feeling of being. J Conscious Stud 12(8–10):43–60

Ratcliffe M (2008a) Feelings of being: phenomenology, psychiatry and the sense of reality. Oxford University Press, Oxford

Ratcliffe M (2008b) Touch and situatedness. Int J Philos Stud Int J Philos Stud 16:299–322

Sass LA (2004) Affectivity in schizophrenia: a phenomenological view. In: Zahavi D (ed) Hidden resources: classical perspectives on subjectivity. Imprint Academic, Exeter, pp 127–147

Solomon RC (1993) The passions: emotions and the meaning of life, revisedth edn. Hackett, Cambridge

Solomon RC (2003) Not passion's slave: emotions and choice. Oxford University Press, Oxford

Solomon RC (ed) (2004a) Thinking about feeling: contemporary philosophers on emotions. Oxford University Press, Oxford

Solomon RC (2004b) On the passivity of the passions. In: Manstead SR, Frijda N, Fischer A (eds) Feelings and emotions: the Amsterdam symposium. Cambridge University Press, Cambridge, pp 11–29

Solomon RC (2007) True to our feelings: what our emotions are really telling us. Oxford University Press, Oxford

Strasser S (1977) Phenomenology of feeling: an essay on the phenomena of the heart. Duquesne University Press, Pittsburgh

Phenomenology, Imagination and Interdisciplinary Research

Julia Jansen

Introduction: Staking Out the Field

The concept of imagination is notoriously ambiguous.[1] Thus one must be cautious not to use 'imagination' as a placeholder for diverse phenomena and processes that perhaps have not much more in common than that they are difficult to assign to some other, better defined domain, such as perception, conceptual thought, or artistic production. However, this challenge also comes with great opportunities: the fecundity and openness of 'imagination' appeal to researchers from different disciplines with different approaches and questions, and it draws together fields of enquiry that are initially considered far apart. Hence, arguably, the field of imagination is particularly poised for interdisciplinary enquiry. In the section on Imagination in Interdisciplinary Research, I will talk about some of the issues that have already entered that field of interdisciplinary inquiry.

This field becomes considerably larger if we also use the term 'imagination' for basic activities which go beyond the mere processing of perceptual data but are still considered integral to perception (e.g. because they occur in the absence of perceptual stimuli). In this respect, Hume's and Kant's accounts in particular are still reflected in contemporary research, albeit in ways which are not always explicit (Lohmar 1998). I will say something about this in the section on Imagination in Interdisciplinary Research too.

Before I turn to the matter of interdisciplinary research, however, I will first, in a section on Imagination in Phenomenology, sketch a general phenomenological position on imagination. I will mainly focus on Husserl's account of phenomenology because it provides a solid reference point for understanding the context from which phenomenological contributions to interdisciplinary research on imagination are put forward.

J. Jansen
Philosophy, University College
Cork, Ireland

[1] Stevenson (2003) identifies thirteen different ways in which imagination is taken up in ordinary language and academic research.

The bulk of Husserl's investigations have only just been translated into English (Husserl 2005) and even in the German original they have only been available since the publication of Husserl's lecture notes on "Phantasy, Image Consciousness, and Memory" in 1980 (Husserl 1980). This has meant that while many of Husserl's observations on imagination had already been known – mostly via Sartre (2004) and, in the analytic tradition, via Warnock (1976) and more recently McGinn (2004) – very few readers would have been aware of them as *his*. Thus it is also in order to rectify this situation somewhat that I will give considerable space to Husserl's analyses.

Most phenomenological contributions to interdisciplinary research, however, are not made by interpreting specific phenomenological texts or authors (although they might be cited in support of particular claims) but by approaching issues in philosophy of mind, psychology, cognitive science and the neurosciences from a generally speaking 'phenomenological' perspective. In this sense, current interdisciplinary work is arguable much closer to the original experimental spirit of phenomenology than any exegesis of textual sources can ever be.

Imagination in Phenomenology

In phenomenology, the concept of imagination has always played a prominent role. It was heralded by Husserl as the 'vital element' of phenomenology (Husserl 1983: 160); appropriated by Heidegger as the ecstatic nature of Dasein[2]; identified by Sartre as "an essential and transcendental condition of consciousness" (Sartre 2004: 188). It has been central to the work of philosophers as diverse as Bachelard (2005),[3] Ricoeur (1977),[4] Castoriadis (1994, 1998).[5] Casey (1976) and others (most recently Marc Richir 2004).

Husserl describes imagining[6] as an act of intuitively (i.e. quasi-perceptually), experiencing something in the mode of 'inactuality' or 'irreality.' He thus distinguishes imagining from a mere supposing or a 'thinking of' (which are intuitively 'empty' acts), but also from remembering (which involves belief in the past reality of the remembered), from expecting (which involves belief in the future reality of the expected), and from perceiving (which involves belief in the present reality of the perceived). The essential differences between imagining and those other acts

---

[2] For an insightful account of Heidegger's engagement with imagination see Elliot (2005).

[3] Bachelard combines an interest in the creative potential of imagination with an ethical and metaphysical committment to imagination as a principle of freedom and transcendence (Kaplan 1972).

[4] Ricoeur's hermeneutical phenomenology approaches imagination in its interconnection not with perception but with language.

[5] Castoriadis explored the political power and effectiveness of a radical 'imaginary.'

[6] Husserl's term is 'phantasy' (Phantasie) (Husserl 2005). In order to make this article more readible but also in order to preserve the connections to other discourses, I decided here to use the term 'imagination' and its derivatives 'imagining,' 'imagined' etc. instead.

are thus not differences amongst their *contents* (nor, as Hume had it, of their differing degrees of vivacity or intensity) but instead differences in the way (or 'form') in which they are experienced (Husserl 1984: 624). For example, whereas in perception "the object appears to us, so to speak, 'in person,' as itself present," in imagination the object appears as represented or possible; "it is as though it were there, but only as though." (Husserl 2005: 18 (16)).

According to Husserl, imagination shares this 'non-positing' character, i.e. the lack of belief in the existence of its object, with what Husserl calls 'picture-consciousness.'[7] This inclines us to think of imagination in analogy with picture-consciousness, that is, with 'seeing something in a picture', hence in terms of 'mental images' that are analogous to pictures. However, as Husserl tries to show, imagining is in a crucial respect very unlike, even essentially different from, 'seeing something in a picture.'[8]

When I look at a picture and see something 'in it,' the object I experience involves three distinct moments: (1) the *physical* picture, i.e., the canvas painted and framed, the patches of colour distributed on the canvas, etc., (2) the *picture-object*, i.e., the image which *appears* through a certain distribution of colours and shapes, and (3) the *picture-subject*, i.e., what is *depicted* or represented by the image.[9] The crucial difference between imagination and picture-consciousness is, according to Husserl, that picture-consciousness requires the perception of an actual picture, while imagination does not.[10]

> If our imagination playfully occupies itself with angels and devils, dwarfs and water nymphs ( ), then the appearing objectivities are not taken as picture objects, as mere representatives, analogues, pictures of other objectivities ( ). The word 'imagination,' the talk of 'mental

---

[7] Brough translates 'Bildbewusstsein' as 'image consciousness' (Husserl 2005). In order to make it more obvious that Husserl refers to pictures (and not to mental images) I will use 'picture consciousness' instead.

[8] This is perhaps the most important difference between Husserl's and Sartre's accounts. See Stawarska (2005). In this paper I can only allude to the aspects of picture-consciousness that are directly relevant to this distinction. For detailed discussions see Brough (1992, 2005), Marbach (1993), Volonté (1995), Lotz (2007).

[9] For example, when we look at a picture and say "this looks just like her!" we do not mean the physical picture (which looks like other physical pictures rather than like a real person), but we mean the picture object, i.e., the image that appears in the picture. On the other hand, the picture object clearly is not the depicted real 'her' who is probably of a different size and color, three-dimensional, moving, etc. (Husserl 1980: 121f. (112)).

[10] It "seems most appropriate to speak of 'pictoriality,' of 'pictorial apprehension' only in cases in which a picture, which for its part first functions as a representing object for something depicted, actually appears. Hence in the case of simple imagination, in which this does not occur (however great the temptation to assume that the situation is the same), it is best to use a different term" (Husserl 2005: 94 (87)). – Initially, Husserl thinks of imagination in pictorial terms. My summary is based on Husserl's mature account. For the considerable changes he made to his earlier position see Jansen (2005).

images' and so forth, ought not mislead us here any more than the talk of 'perceptual images' does in the case of perception (Husserl 2005: 92 (85); translation slightly modified).

Just as one does not, in perception, apprehend a perceptual representation as an image of the perceived object, one does not, in imagination, apprehend a 'mental image' that represents the imagined object. In perception as well as in imagination "the intention aims at the thing itself" (Husserl 2005: 192 (161)). Husserl thus identifies a parallelism between imagination and perception (not between imagination and picture-consciousness). Consequently, he vehemently rejects any 'image-theory' of imagination that would use a "crude talk of internal images (as opposed to external objects)" (Husserl 1985: 437).

The parallelism between imagination and perception underscores the characterisation of imagination as *quasi*-perception. Both intentional acts constitute an object and let it appear; they both have the same intentional structure and are subject to the same spatial and temporal articulation (imagining is an imagining in a *quasi*-here-and-now) (Husserl 1973: 169ff.). In both acts I am also bodily present and thereby have a certain perspective on the intended object, which shows the same horizonal structures in both cases (Marbach 1993: 77).[11] In short, imagination brings to bear the phenomenal aspects of its objects not by conjuring up mental images, which would represent those objects, but by *simulating*[12] experiences of that object.

This simulation can be described phenomenologically in noetic as well in noematic analysis.[13] Noetic analysis shows how, in imagining, a type of experience is reproduced (perceiving visually, perceiving haptically, seeing something in a picture, etc.) while, noematic analysis shows how, at the same time, an object is made present, or represented (Husserl 1969: 128). Strictly speaking, then, in Husserl's view, imagining requires the reproduction of an experience. Or rather, it requires the implication of a possible experience: a simulation.[14]

Imagining is thus essentially different from supposing, which, by contrast, does not imply the simulation of any experiences of the thus supposed objects or states of affairs. It is because we are, in imagination, aware of simulating the experience as well as the object that we notice a "peculiar mediacy" (Husserl 1959: 116), *not* because a doubling up of objects occurs (mental image *plus* imagined object).

Of course, while I'm imagining something, I also actually experience something. For example, while I am imagining (making present) a beautiful beach and imagining (simulating) *possible* ways of experiencing it, namely visually (seeing the white

---

[11] This holds for all sensory modes: visual, tactile, auditory, olfactory, gustatory.

[12] Husserl does not use the term 'simulation' but speaks instead of 'quasi-experience'.

[13] Noetic analyses describe the experience of imagining; noematic analyses describe the imagined object.

[14] Husserl's use of the notion of implication (instead of the now common 'simulation') only highlights the fact that the 'reproduced' act is not actually performed, the 'reproduced' experience not actually experienced, but only 'implied' as a *possible* experience of the imagined object (Marbach: 1993: 61f).

sand, the blue water, etc.), haptically (feeling the sand running through my fingers), etc., I still, unfortunately, perceive my actual surroundings (the desk I'm sitting at, the rain I can hear lashing on the windows), and I am also actually experiencing my imagining being on the beach. When I lose this anchoring in the actual situation I am not imagining but hallucinating or dreaming (Marbach 1993: 83–85).

Furthermore, the phenomenological distinction between *noesis* and *noema* enables an account of the self-ascription of imaginings. That I experience my imaginings as *my* imaginings, as belonging to the one *I* who is now also perceiving, is anything but trivial since the imaginary world is neither subject to 'the legislation of reason', nor constrained by actual space or time (Husserl 2005: 214f. (178)). *Noematically*, in other words, that which I imagine is entirely independent from the constraints of my actual world of experience. There is no necessary noematic link between what we perceive and what we imagine. Noetically, however, our imaginings, perceptions, memories and so forth are united in our one consciousness by our living through them, or experiencing them (Husserl 1975: 175f).

In summary, according to Husserl, imagining lacks belief and reality (it is 'non-positing', 'inactual' or 'irreal') and is in this sense a 'neutralized' representation.[15] It involves a *quasi*-performance or simulation of experiences, such as perceiving, judging, feeling, etc. (Husserl 2005: texts 15, 18a). These experiences are, Husserl observes, implied as possible experiences of the imagined objects. As he repeatedly emphasizes, however, we are not aware of the simulation of such possible experiences as mental representations, but we enact them in our experiences of imagined objects. Any experience can be thus simulated I may imagine myself perceiving, judging, thinking, feeling something and so forth, that is, the imaginative modification is universal.[16]

Husserl advances his theory of imagination not only against contemporary versions of 'image theory' but also against the Kantian notion of a transcendental imagination. Unlike Heidegger, Husserl condemns the 'transcendental imagination' as an unnecessary by-product of Kant's faculty psychology. Especially in his early works, he straightforwardly dismisses Kant's notion as 'untenable' on the grounds that it lacks phenomenological evidence. However, Husserl revises his judgment when he begins his work on 'genetic constitution,' i.e., on constitutive syntheses which Husserl identifies as tacitly effective in perceptual experience. By the time Husserl is working on his Paris lectures (later published in the *Crisis*) he speaks of Kant's "great discovery" of the "twofold operation of the understanding." Husserl

---

[15] Similarly Sartre distinguishes between the positing of perception, which posits its object as existing, and the positing of imagination, which posits its object as nonexistent, absent, existing elsewhere *or* neutralised (Sartre 2004: 12).

[16] Given that Husserl describes imagining as involving the suspension of belief, it remains unclear whether it is possible to simulate the experience of believing. See the section on 'imagination and belief' below for a brief discussion of this issue.

thus extends the meaning of understanding (or rationality) so that it comprises both operations: the personal, explicit, or 'active' syntheses of cognition and judgment; and the sub-personal, tacit, or 'passive' syntheses that configure meaning in the intuitively given life-world (Husserl 1970:103 f.).

Hence Kant's distinction between intellect and transcendental imagination is transformed by Husserl into the distinction between 'active synthesis' and 'passive syntheses.' Importantly, however, for Husserl, passive syntheses are not regulated by concepts, let alone a priori ones (as is Kant's transcendental synthesis of imagination). On the contrary, meaning is generated 'bottom up', so to speak, through the passive syntheses, which are sub-personal perceptual, pre-predicative, pre-reflective and pre-linguistic (Steinbock 2001: xli; Husserl 2001). Thus, Husserl addresses the issue raised by Kant's notion of transcendental imagination but does so by integrating its functions into the complex system of perception itself and thus by making them constitutive of the very understanding that Kant thought was in control of them.

Thinking of phenomenological descriptions in relation to the Kantian notion of transcendental imagination highlights important aspects of a generally speaking phenomenological account of imagination[17]: Although it is discovered in reflection that perception indeed always already exceeds the mere processing of sense data, this is typically not considered evidence for any 'imaginative' activity. On the contrary, it is generally held amongst phenomenologists that perception itself, even in its most elementary moments, "arouses the expectation of more than it contains, and … is therefore already charged with a *meaning*" (Merleau-Ponty 1962: 4). In perception, as Sartre said, "I always *perceive more and otherwise* than I *see*" (Sartre 2004:120). Imagination and perception are therefore considered distinct; "far from being two elementary psychic factors of similar quality and that simply enter into different combinations (they) represent the two great irreducible attitudes of consciousness" (Sartre 2004: 120). He passive or prereflective syntheses which phenomenologists identify as constitutive of experience are hence not, as is Kauf's transcendental synthesis of imagination, understood in terms of a 'top down' process of experience. Rather, phenomenologists attempt to describe the complex ways in which our concepts are grounded in perception and arise through processes of abstraction and formalisation in a "genesis of meaning" (Merleau-Ponty 1962: xix; Husserl 2001; Husserl 1970).

Imagination in Interdisciplinary Research

There has been a relatively recent resurgence of and a more concerted interest in imagination in analytic philosophy of mind. In a climate in which imagination was often considered an expression of old-fashioned idealism or of a misguided preoccupation of continental philosophy, Mary Warnock (1976) took an important step towards rehabilitating it as a serious object of philosophical analysis. Kendall

---

[17]These two aspects are only meant as typical, not as necessary, features of a phenomenologica approach. My rendering is mainly based on Husserl's, Sartre's and Merleau-Ponty's writings.

Walton's later proposal to comprehend children's play, fiction and other artistic productions as different forms of regulated imaginary pretence (Walton 1990) has been immensely influential in debates on pretence, fiction, and emotional responses to fiction; but it also established imagination as an indispensable element of such debates (Currie 1990, 1995, 1997; Meskin and Weinberg 2003). More recently, imagination has entered into wider epistemological and ethical investigations regarding counterfactuals, conceivability, belief and supposition, action theory, 'mind reading,' and creativity (Currie and Ravenscroft 2002; Gendler and Hawthorne 2003; Gendler 2003; Byrne 2004; Nichols 2006; Nettle 2001). This work has begun to cross over into psychology and psychopathology, especially with respect to the role of imagination in autism and schizophrenia, which are both considered 'pathologies of the imagination' (Currie and Ravenscroft 2002; Phillips and Morley 2003). Since phenomenologists, such as Gallagher (2005, 2004), Zahavi (2004, with Parnas 2003, 2001), Casey (2003), Fuchs (2005), Ratcliffe (2006, 2008), are engaging in the very same debates, interdisciplinary imagination research has also brought opportunities for intra-disciplinary discussion.

Until very recently, it seemed inconceivable that the cognitive sciences would (or could) ever pursue the explanation of something as elusive and capricious, but also as subjective and 'mentalistic' as imagination. During the long-lasting reign of behaviourism,[18] imagination seemed pushed, once and for all, into "the outer darkness of intellectual irrelevance" (Morley 2005: 117). However, the so-called 'cognitive turn,' which turned psychology on its head in the 1970s, triggered new interest in imagination and thus brought relief from the 'iconophobia' (Thomas 2007) of the earlier days. Current research on imagination, its neurological manifestations and its psychological (normal as well as pathological) effects is flourishing (Chalmers and Bourget 2007).

In what follows I will briefly outline some of those aspects of imagination research that either have already been taken up in interdisciplinary debate or obviously lend themselves to it. These are: (a) mental imagery; (b) 'mind reading'; (c) imagination and belief; (d) imagination as 'ingredient of perception'; and (e) imagination and aesthetics. Since the issue of mental imagery is, in ways that will become clear below, fundamental to phenomenological views on other matters, I will discuss it in more detail than the others.

In what follows, I will address each issue in turn. Where phenomenological approaches have not been advanced (or where I am not aware of them), I will try to outline possible avenues for such contributions.

(a) *Mental Imagery*. Does imagination or mental imagery imply the existence of mental images? In contemporary research, this question forms the contentious basis of what's known as the 'imagery debate' (Block 1981; Tye 1991). Three principal answers have been given to this question:

(1) *Yes. Mental imagery involves representations that are 'image like.'* In line with a long philosophical tradition, Hannay (1971, 1973), Kosslyn (1980,

---

[18] "I believe we can write a psychology (…) and (…) never use the terms consciousness, mental states, mind, content, introspectively verifiable, imagery, and the like" (Watson 1913).

1994), Tye (1988, 1991), Cohen (1996) defend pictorial accounts of mental imagery on the grounds of psychological and neurological experiments involving, for example, 'mental rotation' (Shepard with Cooper 1982, with Metzler 1971) and 'mental scanning' (Kosslyn 1978) tasks. Differences in the speed and ease with which subjects were able to rotate and scan imagined objects have been found to correlate with the spatial relations of the imagined objects. For example, subjects were able to shift attention from one point to another point on an imagined map faster if the two points were closer together on the corresponding actual map. This has been taken to suggest that mental imagery involves certain spatial features which correspond to, or 'function like' the spatial features of the object represented. Recent neurological research suggests that imagining activates corresponding perceptual visual and motor areas has been taken to lend new support for this thesis (Farah 1988, 1989; Kosslyn et al. 1993) although there also is evidence that this is not always the case (Bértolo 2005).

(2) *No. Mental Imagery involves representations that are not imagistic or pictorial but descriptional.* Objections against the pictorial model include the claim that mental imagery is too indeterminate to represent pictorially (or even 'quasi-pictorially') (Fodor 1975) and the contention that imagery is more like a description than like an image (Dennett 1969; Pylyshyn 1973, 1981). Connected to this position are at least two more general positions: one, that any image theory inevitably leads into the homunculus fallacy; two, that mental imagery is not self-contained but that it depends on background knowledge and tacit conceptual processes (*cognitive penetrability*). Pylyshyn puts it like this: "there is much more to what your mental image does and what it 'looks like' than meets the eye – even the 'mind's eye'" (Pylyshyn 2003a: 6.1). As a result of this view, the distinction between 'imagining' and 'imagining that' disappears, at least on the sub-personal level. It might remain as a phenomenal difference but, according to descriptionalists, this might be more pronounced in experimental conditions in which 'implied task demands' direct the imagining excercises in particular ways. The same objections that were raised against the earlier psychological version also have been directed at its neurological heir. Moreover, as Pylyshyn points out, the explanatory value of recent formulations of pictorialism in which mental images are said to be only 'functionally' pictorial and 'digitized' is questionable (Pylyshyn 2003b).

One of the main objections against the descriptional view is that it fails to account for neurological evidence indicating that neural visual and motor processes active in imagery largely overlap with those active in perception. However, it has been emphasized that this does not necessarily support the pictorial view (Bartolomeo 2002).

(3) *No, because we have no evidence for the existence of any mental representations, pictorial or descriptional.* While the 'imagery debate' continues, there is a growing body of alternative accounts of perception as enactive, embodied and situated that is supported by neurological evidence (Varela et al. 1991; Hurley 1998; Gallagher 2005; Noë 2004; Noë and O'Regan 2001; Clark 1997). This

has had important implications for conceptions of imagination (Hurley 2006; Jeannerod, 1994, 1995, 1997, 2001; Bartolomeo 2002). In line with Husserl's original observations, these accounts corroborate the view that although we have plenty of phenomenological evidence for imagining and mental imagery we have no phenomenological evidence for the existence of mental representations. When we imagine something we experience, in an imaginative way, *what* we imagine, not an 'image' *of* it; just as when we perceive something we experience, in a perceptual way, *what* we perceive, not an 'image' *of* it.[19]

Sartre (2004), Wittgenstein (2001), and Ryle (2002) are usually credited for having advanced some of the strongest and earliest challenges to any such theory of mental representation.[20] In recent years, McGinn (2004), who largely draws on Sartre's writings, has re-asserted a, broadly speaking, phenomenological critique of representationalism. The principle objection against what Sartre calls the 'illusion of immanence' of image theory, the equivalent to Dennett's 'Cartesian theatre' (Sartre 2004; Dennett 1991), is succinctly put by Slezak: what "these doctrines have in common is the mistake of assuming that we apprehend our mental states rather than just *have* them" (Slezak 2002).

In conjunction with the situated and embodied aspects of enactive accounts of perception, this view has led to a theory of imagination as *enacted and embodied simulation* (Thompson 2007b; Thomas 2007; Currie and Ravenscroft 2002).[21] Imagination is thus no longer understood as "an experience in which we seem to see or have a mental picture" but rather as "the activity of mentally representing an object or a scene by way of mentally enacting or entertaining a possible perceptual experience of that object or scene" (Thompson 2007a: 143). As a result, researchers now speak not only of sensory imagination (visual, audio, olfactory, gustatory, haptic) but also of motor imagination, i.e., imagination of *action* (Jeannerod 1994, 1995, 1997, 2001).[22]

(b) *Mindreading.* We often use the metaphor of 'reading someone's mind' when we describe how we figure out what another person believes or feels. Our capacity to make sense of others and their behaviour is generally considered "a prerequisite for normal social interaction" (Frith and Happé 1999:2; Harris 2000). In current

---

[19] Note in this context that Pylyshyn's description of the three levels involved in the explanation of mental imagery corresponds closely with the three moment of picture-consciousness outlined by Husserl: "At the first level we can ask about the content, or what the representation represents – what it is about. (…) At the second level of analysis, we can inquire about the form of the representation, the system of codes by which mental objects can represent aspects of the world. (…) The third level of analysis of mental representations is concerned with how representations are realized in biological tissue or implemented in hardware" (Pylyshyn 2003a).

[20] Not only Sartre but also Ryle and Wittgenstein were familiar with Husserl's work, at least to some extent.

[21] This idea of enacted and embodied simulation, which we also find in Husserl (see above), is not to be confused with the notion of simulation employed in so-called 'simulation theory,' which is thought of as a mental representation rather than an embodied enactment (see below).

[22] The issue of visual imagination dominates debates; investigations of other modes are rare. See, for example, Reisberg (1992).

research the term comprises investigations into our abilities to 'read' other minds as our own (Nichols and Stich 2003).

Recent neuroscientific studies with persons suffering from Autism Spectrum Disorders and schizophrenia have rekindled psychological and philosophical interest. Discussions are currently dominated by two prominent competing theories about the abilities and processes involved in 'mind reading': 'theory theory' (Baron-Cohen 1989, 1995; Leslie 1991; Gopnik and Meltzoff 1997; Carruthers and Smith 1996) and 'simulation theory' (Goldmann 1989; Gordon 1986, 1995; Heal 1986, 1998a, b). Generally speaking, 'theory theory' assumes that we take a (folk-) theoretical stance by means of which we infer what the other believes or feels, while 'simulation theory assumes that we simulate what we would experience if we were in the other's situation. In short, 'theory theory' says we *think* about other people's mental states while 'simulation theory' tells us that we *imagine* them (Currie and Ravenscroft 2002; Zahavi and Parnas 2003). Simulation seemed to receive additional support from neurological evidence, when Giacomo Rizzolatti, with Vittorio Gallese and other members of their research team discovered the activity of mirror neurons[23] (Fogassi et al. 1998; Gallese and Goldmann 1998).

In opposition to the representationalist approach of both 'theory theory' and 'simulation theory' and in line with enactive and phenomenological views on perception, Gallagher defends the claim that "in most intersubjective situations we have a direct understanding of another person's intentions because their intentions are explicitly expressed in their embodied actions, and mirrored in our own capabilities for action" (Gallagher 2005: 224). This view does not exclude that in some situations, in which we find it perhaps more difficult to make sense of someone, we *do* use either a theoretical stance or empathetic imagination. As Zahavi and Parnas have pointed out: "the crucial question is not whether we can predict and explain the behaviour of others, and if so, how that happens, but rather whether such prediction and explanation constitute the primary and most fundamental form of intersubjectivity" (Zahavi and Parnas 2003).

Especially, if we think of imagination as *enacted and embodied* simulation (see above) and not as the possession of a mental state; and if we think of the theoretical stance as an assessment of behaviour and not of mental states; this does not imply the representationalist model criticised by phenomenologists (Gallagher, 2005; Zahavi 2004). It also does not contradict neurological findings, which precisely demonstrate an embodied enactment, not the presence of a mental presentation (Gallagher 2005; Hurley 2006; Lohmar 2006, 2008). In this way then, we can integrate evidence that autism at least often includes imaginative disorders (Currie and Ravenscroft 2002) into a more general phenomenological account of autism. For example, one could investigate the ways in which difficulties in imaginatively transposing oneself into a different situation or perspective are related to an impaired imaginary of the self which can accompany more basic forms of self-awareness just as self-knowledge can (Zahavi and Parnas 2003: 67; Raffman, 1999). This might also make phenomenological

---

[23] Mirror neurons are neurons in the pre-motor cortex that display the same patterns of activity when an action is observed as they display when an action is performed. In that sense, they are said to 'mirror' in the observer the neurological activity present in the performer of the action.

approaches to empathy (Stein 1989) relevant for contemporary research in social neuroscience (Decety and Grèzes 2006; Decety and Hodges 2006).

In a similar way, current phenomenological contributions to interdisciplinary research on schizophrenia (Danion and Huron 2007; Gallagher 2004; Parnas 2004, with Handest 2003, with Parnas and Sass 2002; Parnas and Sass 2006; Depraz 2003; Zahavi 2001) can be augmented by an exploration of the 'imaginative disorders' involved both in schizophrenic delusion as well as schizophrenic hallucination (Casey 2003). All these enquiries demand careful attention to the ways in which our awareness of self, of others and of objects is *enacted and embodied* (Fuchs 2005).

(c) *Imagination and Belief*. Investigations into the relation between imagination and belief obviously arise from a notion of imagination as 'make believe' or pretence. These debates are gathering additional momentum from psychological and neuroscientific research into developmental and pathological issues of pretence, for example in autistic children (Harris 2000; Hurley with Chater 2005; Frith and Happé 1999; Currie and Ravenscroft 2002).

However, there is also a perhaps more fundamental question about differences between imagining something (without believing it) and supposing it (without believing it). This question becomes more complex again considering the view that one can imagine beliefs, just as one can imagine experiences and desires, and that supposing something *is* imagining a belief (Currie and Ravenscroft 2002). Phenomenological analysis, however, casts doubt on such a conception. We do not have to simulate the experience of believing in a state of affairs in order to suppose that state of affairs. In fact, a 'mere' supposition is precisely characterized by a lack of belief. Rather, in supposing something, or in 'imagining that …' we consider a state of affairs possible, while suspending the question of whether we believe in it or not. Neither the *noetic* aspect of imagining (the simulation of an experience) nor the *noematic* aspect of imagining (the simulation of the object) need be present in supposing.

Further, it is questionable whether we can simulate belief. Belief is a mode in which we intend something, e.g., we can entertain a thought in the mode of belief, or doubt, or under the suspension of belief-we can perceive with belief (as we usually do), but we can also doubt what we perceive (when, for example, we are aware of a perceptual illusion). To talk about 'belief-like imagining' hence implies that we can simulate that particular mode of experiencing something *with belief*. It is unclear whether it is possible to do that but it seems pretty clear that this is not what we do in supposing or when we 'imagine that …' Thus it seems that if, in line with ordinary language, we still want to refer to such acts as instances of imagining, then at least we have to note this essential difference between them (O'Brien 2005).

The fact that in experience cases of mere supposing and imagining are not always clearly distinguishable is a good reason for considering those issues as related. Recent work on counterfactual thinking, which includes both activities, has shown the great extent to which many rational operations depend on them (Byrne 2004). Moreover, research on 'imaginative contagion' has shown that either can have great impacts on our (actual) beliefs (Gendler 2007).

(d) *Imagination as 'Ingredient of Perception.'* Hume's and Kant's idea that imagination has a necessary (Hume) or transcendental (Kant) function in perception is still part of contemporary discourse, whether imagination is (Johnson 1987, with Lakoff 1999) or is not (Prinz 2002, McDowell 1994) explicitly referred to. The underlying idea is that we do not just 'see' what there is but that we also import or project certain elements into perception. In other words, there is evidence of a gap between present perceptual stimuli and what is perceptually experienced as present. Recent research on 'change blindness' (the failure to perceive even great changes in a perceived scene), 'inattentional blindness' (the failure to perceive events outside one's attentional focus), and 'filling-in' or 'perceptual completion' (the seeing of a figure as complete although parts of it are outside the visual field) has stirred a debate on the so-called 'grand illusion' of perception (Noë 2002a). Contributions to this debate from the 'enactive' and 'embodied' perspective make use of the phenomenological evidence already invoked by Husserl, Merleau-Ponty and Sartre to highlight the differences between perception and imagination and thus to reject the hypothesis of a 'grand illusion' (Noë 2002b; Thompson et al. 1999). However, paying special attention to the differences – within perception – between those elements that can be explained with reference to present stimuli and those that cannot surely amounts to important research into the nature of perception and does not require representationalism (O'Connor and Aardema 2005; Lohmar 2008; Lennon 2009 forthcoming).

(e) *Imagination and Aesthetics.* I mentioned above the strong interest in 'make believe', fiction and emotions manifest in recent discussion in aesthetics. Traditionally, imagination has played a great role in conceptions of aesthetic experience as well. However, while cognitive and neuro-scientists have turned towards aesthetic experience and, in particular, to our experience of beauty (Kawabata and Zeki 2004; Zeki 1999a,b, 2002; Romano 2002; Blood and Zatorre 2001; Ramachandran and Hirstein 1999; Solso 1996), there has been little phenomenological response to the emerging discipline of neuroesthetics.[24] Perhaps aesthetic experience is considered too complex and too phenomenologically under-researched for focused interdisciplinary debate. And there are reasons to be sceptical about initial interpretations of neurological findings (Seeley 2006; Jansen 2006; Ione 2003). However, this only calls for more phenomenological clarification of aesthetic experience and its imaginative dimension, which in recent years has perhaps not received as much attention as it should.

---

[24]Lead by Semir Zeki, researchers at University College London and University of California at Berkeley founded the Institute of Neuroesthetics in 2002. For further information, see http://www.neuroesthetics.org/index.html.

Conclusion

Whereas each of these issues constitutes an individual research area that has largely been pursued independently from all the others, it is also possible to regard these areas as moments of an emerging integrated field of inquiry profiting – by means of processes of mutual constraint and instruction (Gallagher 1997) – from cross-fertilization across disparate disciplines, such as philosophy, cognitive science, neuroscience, psychology, psychopathology and psychiatry. This perspective keeps open a space in which further research will show whether and how the many different senses of imagination can be related phenomenologically, conceptually, psychologically, or neurologically. This, I take it, is the best reason to approach imagination in phenomenology and interdisciplinary research with a 'holistic stance' and – if only now and again, when one looks up from one's own specialized work – consider it as a multifaceted but integrated field of research.

References

Bacheland G (2005) The poetics of space (trans: Jolas M). Beacon Press, Boston MA

Baron-Cohen S (1989) The autistic child's theory of mind. J Child Psychol Psychiatry 30:285–298

Baron-Cohen S (1995) Mindblindness: an essay on autism and theory of mind. MIT Press, Cambridge, MA

Bartolomeo P (2002) The relationship between visual perception and visual mental imagery: a reappraisal of the neuropsychological evidence. Cortex 38:357–378

Bértolo H (2005) Visual imagery without visual perception? Psicológica 26:173–188

Block N (ed) (1981) Imagery. MIT Press, Cambridge, MA

Blood A, Zatorre R (2001) Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. PNAS 98(20):11818–11823

Brough J (1992) Some Husserlian comments on depiction and art. Am Cathol Philos Q LXVI(2): 241–259

Brough J (2005) Translator's introduction. In: Husserl E, Phantasy, image consciousness, and memory (1898–1925). Springer, Dordrecht, pp XXIX–LXVIII

Byrne R (2004) The rational imagination. MIT Press, Cambridge, MA

Carruthers P, Peter KS (1996) Introduction. In: Carruthers P, Smith PK (eds) Theories of theories of mind. Cambridge University Press, Cambridge

Casey ES (1976) Imagining: a phenomenological study. Indiana University Press, Bloomington, IN

Casey ES (2003) Imagination, fantasy, hallucination, and memory. In: Philips J, Morley J (eds) Imagination and its pathologies. MIT Press, Cambridge

Castoriadis C (1994) Radical imagination and the social instituting imaginary. In: Robinson G, Rundell J (eds) Rethinking imagination: culture and creativity. Routledge, New York

Castoriadis C (1998) The imaginary institution of society (trans: Blamey K). MIT Press, Cambridge, MA

Chalmers D, Bourget D (eds) (2007) Mindpapers. A bibliography of philosophy of mind and science of consciousness. http://consc.net/mindpapers/.

Clark A (1997) Being there: putting brain, body, and world together again. MIT Press, Cambridge, MA

Cohen J (1996) The imagery debate: A Critical Assessment. J Philos Res 21:149–182

Currie G (1990) The nature of fiction. Cambridge University Press, Cambridge

Currie G (1995) Image and mind. Cambridge University Press, New York

Currie G (1997) The paradox of caring. In: Hjort M, Laver S (eds) Emotion and the arts. Oxford University Press, Oxford, pp 63–77

Currie G, Ravenscroft I (2002) Recreative minds. Oxford University Press, Oxford

Danion J-M, Huron C (2007) Can we study subjective experiences objectively? First-person perspective approaches and impaired subjective states of awareness in schizophrenia? In: Zelazo PD, Moscovitch M, Thompson E (eds) The Cambridge handbook of consciousness. Cambridge University Press, Cambridge, pp 481–498

Decety J, Grèzes J (2006) The power of simulation: imagining one's own and other's behavior. Brain Res 1079:4–14

Decety J, Hodges SD (2006) The social neuroscience of empathy." In: van Lange PAM (ed) Bridging social psychology: benefits of transdisciplinary approaches. Erlbaum, Mahwah, pp 103–109

Dennett D (1969) Content and consciousness. Routledge & Kegan Paul, London

Dennett D (1991) Consciousness explained. Little, Brown, Boston, MA

Depraz N (2003) Putting the epoche into practice: schizophrenic experience as illustrating the phenomenological exploration of consciousness. In: William K, Fulford M, Morris KJ, Sadler JZ, Stanghellini G (eds) Nature and narrative: an introduction to the new philosophy of psychiatry. Oxford University Press, Oxford

Elliot B (2005) Phenomenology and imagination in Husserl and Heidegger. Routledge, New York

Farah M (1988) Is visual imagery really visual? Overlooked evidence from neuropsychology. Psychol Rev 95(3):307–317

Farah M (1989) The neural basis of mental imagery. Trends Neurosci 12:395–399

Fodor JA (1975) The Language of thought. Thomas Crowell, New York

Fogassi L, Gallese V, Fadiga L, Rizzolatti G (1998) Neurons responding to the sight of goal-directed hand/arm actions in the parietal area PF (7b) of the macaque monkey. Soc Neurosci Abstr 24:654

Frith, Uta and Francesca Happé. 1999. "Theory of Mind and Self-Consciousness: What is it like to be autistic?'" Mind and Language 14(1):1–22

Fuchs T (2005) Corporealized and disembodied minds: a phenomenological view of the body in melancholia and schizophrenia. Philos Psychiat Psychol 12(2):95–107

Gallagher S (1997) Mutual enlightenment: recent phenomenology in cognitive science. J Conscious Stud 4(3):195–214

Gallagher S (2004) Agency, ownership, and alien control in schizophrenia. In: Dan Z, Grunbaum T, Parnas J (eds) The structure and development of self-consciousness: interdisciplinary perspectives. John Benjamins, Amsterdam

Gallagher S (2005) How the body shapes the mind. Oxford University Press, Oxford

Gallese V, Goldman A (1998) Mirror neurons and the simulation theory of mind-reading. Trend Cogn Sci 2:493–501

Gendler TS (2003) On the relation between pretense and belief. In: Kienan M, Mcluer Lopes D (eds) Imagination, philosophy, and the arts. Routledge, London, pp 125–141

Gendler TS (2007) Imaginative contagion. Metaphilosophy 37(2):183–203

Gendler TS, Hawthorne J (eds) (2003) Imagination, conceivability and possibility. Oxford University Press, Oxford

Goldmann A (1989) Interpretation psychologized. Mind Lang 4:161–185

Gopnik A, Meltzoff A (1997) Words, thoughts, and theories. MIT Press, Cambridge, MA

Gordon R (1986) Folk psychology as simulation. Mind Lang 1:158–171

Gordon R (1995) Simulation without introspection or inference from me to you. In: Davies M, Stone T (eds) Mental simulation: evaluations and applications. Blackwell, Oxford

Hannay A (1971) Mental images – a defence. Allen & Unwin, London

Hannay A (1973) To see a mental image. Mind 82:161–262

Harris P (2000) The work of the imagination. Blackwell, Oxford

Heal J (1986) Replication and functionalism. In: Butterfield J (ed) Language, mind, and logic. Cambridge University Press, Cambridge, pp 45–59

Heal J (1998a) Co-cognition and off-line simulation: two ways of understanding the simulation approach. Mind Lang 13:477–498

Heal J (1998b) Understanding other minds from the insides. In: O'Hear A (ed) Current issues in philosophy of mind. Cambridge University Press, Cambridge:

Hurley S (1998) Consciousness in action. Harvard University Press, Cambridge, MA

Hurley S (2006) Active perception and perceiving action: the shared circuits hypothesis. In: Gendler TS, Hawthane J (eds) Perceptual experience. Oxford University Press, Oxford, pp 205–259

Hurley S, Chater N (eds) (2005) Perspectives on imitation: from cognitive neuroscience to social science. MIT Press, Cambridge, MA

Husserl E (1959) Erste Philosophie (1923/24). Zweiter Teil: Theorie der phänomenologischen Reduktion (= Husserliana VIII), Boehm R (ed). Martinus Nijhoff, The Hague [my translation]

Husserl E (1969) Zur Phänomenologie des inneren Zeitbewusstseins (1893–1917) (= Husserliana X), Boehm R (ed.) 2nd edition. Martinus Nijhoff, The Hague [my translation]

Husserl E (1970) Crisis of the European sciences and transcendental phenomenology (trans: David C). Northwestern, Evanston

Husserl E (1973) Experience and judgment (trans: Churchill JS, Ameriks K). Northwestern, Evanston

Hasserl E (1975) Log ische Uutersuchungen Erster Band (= Husserliana XIII), Holenstein E (ed.). Martinus Nijhoff The Hague

Husserl E (1980) Phantasie, Bildbewusstsein, Erinnerung. Zur Phänomenologie der anschaulichen Vergegenwärtigungen (1898–1925) (= Husserliana XXIII), Marbach E (ed.). Martinus Nijhoff, The Hague

Husserl E (1983) Ideas pertaining to a pure phenomenological philosophy. First Book (trans: Fred K). Martinus Nijhoff, The Hague

Husserl E (1984) Logische Untersuchungen. Zweiter Band: Untersuchungen zur Phänomenologie und Theorie der Erkenntnis, vol. 2 (= Husserliana: XIX/2), Panzer U (ed). Kluwer Dordrecht [my translation],

Husserl E (1985) Einleitung in die Logikund Eikenttin's theories Vorelesungen 1906/07 (= Husserliana XXIV), Melle U (ed.). Kluwer, Dordrecht

Husserl E (2001) Analysis concerning passive and active syntheses (trans: Steinbock A). Kluwer, Dordrecht

Husserl E (2005) Phantasy, image consciousness, and memory (1898–1925) (trans: Brough J). Springer, Dordrecht [pagination in square brackets follows the German original (Husserl 1980)].

Ione A (2003) Examining Semir Zeki's neural concept of formation and art: Dante, Michelangelo, Wagner. J Conscious Stud 10(2):58–66

Jansen J (2005) On the development of Husserl's transcendental phenomenology of imagination and its use for interdisciplinary research. J Phenomenol Cogn Sci 4(2):121–132

Jansen J (2006) Schnittstelle und Brennpunkt: Das ästhetische Erlebnis als Aufgabe für eine Kooperation von Phänomenologie und Neurowissenschaft. In: Lohmar D, Fonfara D (eds) Interdisziplinäre Perspektiven der Phänomenologie. Springer, Dordrecht, pp 142–163

Jeannerod M (1994) The representing brain: neural correlates of motor intention and imagery. Behav Brain Sci 17:187–245

Jeannerod M (1995) Mental imagery in the motor context. Neuropsychologia 33:1419–1432

Jeannerod M (1997) The cognitive neuroscience of action. Blackwell, Oxford

Jeannerod M (2001) Neural simulation of action: a unifying mechanism for motor cognition. Neuroimage 14:103–109

Johnson M (1987) The body in the mind. University of Chicago Press, Chicago/London

Johnson M, Lakoff G (1999). Philosophy in the flesh: the embodied mind and its challenge to western thought. Basic Books, New York City

Kaplan E (1972) Gaston Bachelard's philosophy of imagination. Philos Phen Research 33(1): 1–24

Kawabata H, Zeki S (2004) Neural correlates of beauty. J Neurophysiol 91:1699–1705

Kosslyn SM (1978) Measuring the visual angle of the mind's eye. Cogn Psychol 10:356–389

Kosslyn SM (1980) Image and mind. MIT Press, Cambridge, MA

Kosslyn SM (1994) Image and brain: the resolution of the imagery debate. MIT Press, Cambridge, MA

Kosslyn SM, Alpert NM, Thompson WL, Maljkovic V, Weise SB, Chabris CF, Hamilton EE, Rauch SL, Buonanno FS (1993) Visual mental imagery activates topographically organized visual cortex: PET investigations. J Cogn Neurosci 5:263–287

Lennon K (2009, forthcoming) Imagination and the imaginary. Routledge, New York/London

Leslie A (1991) The theory of mind impairment in autism: evidence for a modular mechanism of development? In: Whiten A, Natural theories of mind. Blackwell, Oxford, pp 63–78

Lohmar D (1998) Erfahrung und kategoriales Denken. Hume, Kant und Husserl über vorprädikative Erfahrung und prädikative Erkenntnis. Kluwer, Dordrecht

Lohmar D (2006) Mirror neurons and the phenomenology of intersubjectivity. Phenomenol Cogn Sci 5(1):5–16

Lohmar D (2008) Phänomenologie der schwachen Phantasie. Untersuchungen der Psychologie, Cognitive Science, Neurologie und Phänomenologie zur Funktion der Phantasie in der Wahrnehmung. Springer, Dordrecht

Lotz C (2007) Depiction and plastic perception. A critique of husserl's theory of picture consciousness. Cont Philos Rev 40:171–185

Marbach E (1993) Mental representation and consciousness. Kluwer, Dordrecht

McDowell J (1994) Mind and world. Harvard University Press, Cambridge, MA

Merleau-Pondg M (1962) Phenomenology of perception (trans: Smith C). Humanities Press, New York

McGinn C (2004) Mindsight. Harvard University Press, Cambridge, MA

Meskin A, Weinberg J (2003) Emotions, fiction, and cognitive architecture. Br J Aesthet 43(1):18–34

Morley J (2005) Introduction: phenomenology of imagination. Phenomenol Cogn Sci 4:117–120

Nettle D (2001) Strong imagination: madness, creativity and human nature. Oxford University Press, Oxford

Nichols S (ed) (2006) The architecture of the imagination. New essays on pretence, possibility, and fiction. Oxford University Press, Oxford

Nichols S, Stich S (2003) Mindreading: an integrated account of pretence, self-awareness, and understanding of other minds. Oxford University Press, Oxford

Noë A (ed) (2002a) Is the visual world a grand illusion? Imprint Academic, Thorverton/Charlottesville, VA

Noë A (2002b) Is the visual world a grand illusion. In: Noë A (ed) Is the visual world a grand illusion. Imprint Academic, Thorverton/Charlottesville, VA

Noë A (2004) Action in perception. MIT Press, Cambridge, MA

Noë A, Kevin O'Regan J (2001) A sensorimotor approach to vision and visual consciousness. Behav Brain Sci 24(5):939–973

O'Brien L (2005) Imagination and the motivational view of belief. Analysis 65(1):55–62

O'Connor K, Aardema F (2005) The imagination: cognitive, pre-cognitive, and meta-cognitive aspects. Conscious Cogn 14(2):233–256

Parnas J (2004) Belief and pathology of self-awareness: a phenomenological contribution to the classification of delusions. J Conscious Stud 11(10–11):148–161

Parnas J, Handest P (2003) Phenomenology of anomalous self-experience in early schizophrenia. Compr Psychiat 44(2):121–134

Parnas J, Sass L (2002) Self, solipsism, and schizophrenic delusions. Philos Psychiat Psychol 8(2–3):101–120

Parnas J, Sass L (2006) Explaining schizophrenia: the relevance of phenomenology. In: Chung M, Fulford W, Graham G (eds) Reconceiving schizophrenia. Oxford University Press, Oxford, pp 63–96

Phillips J, Morley J (2003) Imagination and its pathologies. MIT Press, Cambridge, MA

Prinz J (2002) Furnishing the mind. MIT Press, Cambridge, MA

Pylyshyn Z (1973) What the mind's eye tells the mind's brain: a critique of mental imagery. Psychol Bul 80:1–25

Pylyshyn Z (1981) The imagery debate: analogue media versus tacit knowledge. Psychol Rev 88:16–45

Pylyshyn Z (2003a) Seeing and visualizing: it's not what you think. MIT Press, Cambridge, MA

Pylyshyn Z (2003b) Return of the mental image: are there pictures in the brain? Trends Cogn Sci 7:113–118

Raffman D (1999) What autism may tell us about self-awareness: a commentary on Frith and Happé. Mind Lang 14(1):23–31

Ramachandran V, Hirstein W (1999) The science of art. J Conscious Stud 6(6–7):15–51

Ratcliffe M (2006) Phenomenology, neuroscience and intersubjectivity. In: Dreyfus H, Mark W (eds) Blackwell companion to phenomenology and existentialism. Blackwell, Oxford, pp 327–343

Ratcliffe M (2008) Feelings of being: phenomenology, psychiatry and the sense of reality. Oxford University Press, Oxford

Reisberg D (ed) (1992) Auditory imagery. Lawrence Erlbaum, Hillsdale, NJ

Richir M (2004) Phantasia, imagination, affectivité. Millon, Bonchamp-Lès-Laval

Ricoeur (1977) The Rule of Metaphor (trans: Czerhy R with K McCoughlin and J Costello). University of Toronto Press, Toronto

Romano J (2002) The mind's eye: neuroscience, synesthesia, and art. Neurol Rev 10(7)

Ryle G (2002) The concept of mind. University of Chicago Press, Chicago, IL. Originally published in 1949

Sartre J-P (2004) The imaginary (trans: Webber J). Routledge, New York

Seeley WP (2006) Naturalizing aesthetics: art and the cognitive neuroscience of vision. J Vis Art Prac. Special Issue: Rediscovering Aesthetics, Halsall F, Jansen J, O'Connor T (eds), 5(3):195–213

Shepard RN, Cooper L (1982) Mental images and their transformations. MIT Press, Cambridge, MA

Shepard RN, Cooper L, with Metzler J (1971) Mental rotation of three-dimensional objects. Science 171:701–703

Slezak P (2002) The imagery debate: Déjà vu all over again? Commentary on Zenon Pylyshyn. Behav Brain Sci 25(2):209–210

Solso R (1996) Cognition and the visual arts. MIT Press, Cambridge, MA

Stawarska B (2005) Defining imagination: Sartre between Husserl and Janet. Phenomenol Cogn Sci 4(2):133–153

Stein E (1989) On the problem of empathy (trans: Stein W). ICS Publications, Washington, DC

Steinbock A (2001) Translator's introduction. In: Husserl E (ed) Analysis concerning passive and active syntheses. Kluwer, Dordrecht, pp xvi–lxv

Stevenson LF (2003) Twelve conceptions of imagination. Br J Aesthet 43(3):238–259

Thomas N (2007) Mental imagery. Stanford encyclopedia of philosophy. http://plato.stanford.edu/entries/mental-imagery

Thompson E (2007a) Look again: phenomenology and mental imagery. Phenomenol Cogn Sci (6):137–170

Thompson E (2007b) Mind in life: biology, phenomenology, and the sciences of mind. Harvard University Press, Cambridge, MA

Thompson E with Noë A, Pessoa L (1999) Perceptual completion: a case study in phenomenology and cognitive science. In: Petitot J, Varela F, Pachoud B, Roy J-M (eds) Naturalizing phenomenology. Stanford University Press, Stanford

Tye M (1988) The picture theory of mental images. Philos Rev (97):497–520

Tye M (1991) The imagery debate. MIT Press, Cambridge, MA

Varela F, Thompson E, Rosch E (1991) The embodied mind: cognitive science and human experience. MIT Press, Cambridge, MA

Volonté P (1995) Husserls Phänomenologie der Imagination. Verlag Karl Alber, München

Walton K (1990) Mimesis as make-believe. Harvard University Press, Boston, MA

Warnok M (1976) Imagination. Faberand Faber, London

Watson JB (1913) Psychology as the behaviorist views it. Psychol Rev 20:158–177

Wittgenstein L (2001) Philosophical investigations, 3rd ed. (trans: Anscombe GEM). Blackwell, Oxford

Zahavi D (2001) Schizophrenia and self-awareness. Philos Psychiatry Psychol 8/4:339–341

Zahavi D (2004) Beyond empathy. Phenomenological approaches to intersubjectivity. In Dermot M, Embree L (eds) Phenomenology. Critical concepts in philosophy, vol II. Routledge, London/New York, 179–200

Zahavi D, Parnas J (2003) Conceptual problems in infantile autism research: why cognitive science needs phenomenology. J Conscious Stud 10(9):53–71

Zeki S (1999a) Inner vision. Oxford University Press, Oxford

Zeki S (1999b) Art and the brain. J Concious Stud 6(6–7):76–96

Zeki S (2002) Neural concept formation and art: Dante, Michelangelo und Wagner. J Conscious Stud 9(3):53–76

The Function of Weak Phantasy in Perception and Thinking

Dieter Lohmar

At first glance phantasy seems to be an unessential part of our conscious life, and if we limit ourselves to a purely empirical view of consciousness, it may be difficult to recognize the powerful and central performance of imagination. Thus phenomenology as a descriptive analysis of the essential traits of our conscious life has to analyze phantasy and its functions. I will delineate some of the ways of analysis and also make clear how we can relate phenomenological and transcendental reasoning to empirical research and cognitive science. My aim is to establish the transcendental function of phantasy in perception and in higher cognitive acts directed at states of affairs and the intentions of others by showing that humans apprehend these cognitive contents phantasmatically.

My first main thesis is that weak phantasmata perform a decisive function in human and animal perception. More precisely, I view these "helping phantasmata" as performing a "transcendental function" of perception. Without such a function of imagination, perception would not be possible. Thus in the first part, I argue for the factuality and necessity of phantasmata in perception, drawing on Kant's remark concerning the existence of weak phantasmata in perception. Then we will see that there are phantasmata accompanying our perception in every field of sensuality. To understand the function of phantasma in perception, it will be necessary to discuss the essential traits of apperception with the help of a type (*Typus*).

The second part frames the function of phantasmata into a general theory of cognition. The last part relates human mental facilities to animal consciousness and cognitive performances. I thereby aim to establish an "inclusive theory of cognition and thinking," trying to overcome the one-sided interpretation of human facilities allegedly based on the use of linguistic concepts. Therefore, we have to understand how animals are able to conceive objects, their properties and relations between them, how they have insights into other minds, cultural life and moral rules on the basis of non-linguistic modes of thinking and cognition

D. Lohmar
University of Cologne,
Philosophisches Seminar

that employ a phantasmatic medium. Additionally, we have to understand how parallel to the linguistic mode of thinking, human beings still continue to think in these non-linguistic modes.

Thus I will try to establish the concept of a "symbolic system of representation" as denoting a general concept of a performance within which our language is only one single case. Yet a general idea is best explained through language: A system of representation should enable us to conceive of states of affairs and events without having an appropriate intuition of them. Humans usually do this using linguistic expressions.

The argument of the second part will run along four theses:

1. Language is one system of representation but we can in principle conceive of alternative systems of representation that would be equally effective. In my view, Husserl's phenomenological theory of meaning leaves open the possibility of systems of representation other than language.[1]
2. There must necessarily be non-linguistic systems of representation in humans and in higher primates. The arguments for this thesis builds upon theories of biological development, the history of human evolution and new insights into the mental abilities of higher primates that are all achieved without language.
3. A non-linguistic system of representation is still functioning in our own consciousness. We use simultaneously different systems of representation, among which are language, gestures, feelings and scenic images. This claim will find support in the phenomenological analysis of the non-linguistic systems functioning in us. It is especially fruitful to investigate the scenic mode of daydreaming as a central form of non-linguistic modes of thinking.
4. It is highly probable that the non-human members of the primate group are able to think by using the same non-linguistic systems of representation as we still do.

My method is that of a phenomenological description, mostly oriented by the guidelines of Husserl. This method allows for a detailed analysis of the activities of the human mind in perception, insight, decision-making, etc. Phenomenological description is a "view from within" which has a different character than third person empirical research. This philosophical groundwork tries to disclose the essential traits of human consciousness and the necessary conditions of human cognition. But phenomenology does not only function as a philosophical groundwork, it also investigates the same object as does empirical psychology, i.e. everyday consciousness and its variants. So there are a lot of overlapping themes and even similarity in the conclusions that allows for a comparison and an attempt to bring both methods into a fruitful interplay.

---

[1] Most of the theses proposed here are also to be found in Lohmar (2008). I would like to express my gratitude to Saulius Geniusas for his kind help with the English text.

Weak Phantasmata in Perception

Weak phantasmata are best characterized as a kind of sketching-in of imagination into the intuitive realm of sensuality. This sketching-in or drawing-in is possible in all fields of sensibility: optical, acoustic, tactual as well as in taste and smell.[2] I call these productions of imagination "phantasmatic self-affection," or simply phantasmata. For example, if you see someone biting into a lemon, something happens within your own taste-field, you somehow feel as if there is something sour in your taste.

Kant was one of the first to describe in detail this seemingly psychological effect in some of his precritical writings.[3] He describes our imagination as being able to draw-in something into our visual sensuality (hear-in, smell-in, etc.). The result of this self-affection is nearly as real as the original sensation. It appears to us like an affection "from outside". Kant's preferred example for this faculty of the mind is lying in one's bed just woken up and looking at some rough wall or curtain at the bedside. Kant states that we are able to see faces, Gestalts or whole scenes in this amorphous material. In this special case of perception, most of the features of the faces perceived are somehow given in sensuality.

Phenomenologically interpreting this kind of givenness, we might say: There are kinds of lines[4] which can be interpreted as part of a Gestalt, face or scene. If we are able to perceive the face in the curtain, then we have to combine the sensual elements (which gain new sense by the foregoing interpretation) forming a representation of the object perceived. Thus we might view this process of synthetic activity as interpretation, mental connection and combination of the given lines to present the thing we see.

But Kant's really astonishing discovery was that there might appear a difference in sensuality when I see the face or when I see the curtain: Some important features or lines of the face appear when I see the face and disappear when I concentrate on the curtain. Kant named these strangely appearing and disappearing traits "chimerical traits". We may be even able to influence this activity of our imagination arbitrarily. When we see the face in the curtain these traits are present, but they vanish if we see the curtain.[5] In Kant's view, this ability is not a sign of mental disease, it occurs in every healthy mind.[6]

---

[2] I have mentioned only the classical five senses, but there are more, for example, the sense of movement (or being moved), cf. Berthoz (2002). We might also regard our feelings as something produced by our imagination (in a very broad sense), because there is no simple causal explanation for our feelings. Nevertheless feelings are felt and experienced by us as something which is unwillingly given to us (like sensuality). Thus feelings may be also regarded as phantasmata.

[3] The descriptions are to be found in Kant (1764) and Kant (1766). Already Aristoteles has a theory of phantasmata, as well as Th. Aquinus.

[4] Husserl would prefer to speak of prominances (Abgehobenheiten).

[5] Cf. Kant (1764), 265 and Kant (1766), 346, Anm.*

[6] Cf. Kant (1764), 264 f. But this ability of the human mind is also the basis for sensual self-deception ("Selbstbetrug in den Sinnen", ibid.). But Kant emphasizes that it is an activity of the mind that "commonly happens and even should happen in healthy humans" (my transl. of "die in gewöhnlicher Weise bei gesunden Menschen geschieht und auch geschehen soll", Kant (1766), 340, cf. also 344).

At first glance, this case of perception seems to be quite exceptional, but Kant views it as an extreme case of normal perception. Even in his *Critique of Pure Reason*, normal perception shares most characteristics with this seemingly extraordinary case. In normal perception we are given only disconnected chaotic parts in sensuality which we have to interpret, combine and connect in our thinking to constitute a representation which can present the object to me. Kant's critical writings hold on to the basic concept of perception presented in this example: The chaotic material of intuitive sensibility together with the conceptual forming, ordering and connecting activity of the mind. Yet it seems that in the *Critique* Kant almost lost sight entirely of the ability of imagination to vividly sketch-in parts of the objects perceived with the help of phantasmata.

Thus we must go further in our analysis beyond Kant. The main reason for this is not that he lost sight of the function of phantasma in perception but that he, following the tradition of the older theological rationalism, uses a concept of reason that belongs exclusively to humans and therefore does not allow for an inclusive theory of perception and thinking.

Let me extend the realm of examples of phantasmatic self-affection to show that it can function in every field of sensuality. The most important class of examples for phantasmatic self-affection arises in normal perception, especially if there is a weakening of sensuality, missing of sensual contents or a complete loss of one field of sensuality. In the case of the lemon: We see it but we do not taste it, so that the sensation of the sourness is sketched-into a more or less empty field of sensuality.

The "inner voice" is usually accompanying our own thinking with something similar to the hearing of our own voice. This inner voice sounds like our own voice as we hear it while speaking. Although everyone knows about this experience, we do not know what function it has. But we are also able to produce phantasmatic voices of other persons: If we read a letter from a good friend or a text of our academic teacher we are able to hear his voice in his individual tone and temperament.

You may also try the following: While you are speaking with a friend on the phone, just close your eyes – by this the visual field will be darkened – and you are able to see the person looking at you while speaking and even gesturing. It is even possible that through the tone of the voice you can learn about his present feelings and then your phantasmatic accompaniment will reveal corresponding facial expression and gestures.

If all fields of sensuality are darkened, as in sleep, then we are able to have phantasmata in most of our sensual fields – we dream. But there are serious shortcomings of dreaming, which makes it not acceptable as a field of phenomenological description: In dreams we are not awake and reflectively self-conscious.

But my central argument in regard to phantasmatic self-affection does not depend on our ability to have all kinds of such phantasmata. I will only presuppose that we are in principle able to produce some of these different kinds. To determine the precise sense of this activity, I will now outline an analysis of the process of perceiving which Husserl understands as an apperception (*Auffassung*) of given sensuality with the help of a type (*Typus*). In this analysis I would like to make clear

the function of the weak "helping phantasmata" we have already become acquainted with in some examples.

I will pursue my analysis in the framework of genetic phenomenology, using the concept and the function of types. A first explanation of a type may be: It is an idea in which a number of partial intentions are collected and bound together so as to belong to a certain kind of object, for example a tree. We not only have types of general things like trees, humans or books but also of singular things, for example my friend Peter. This type of a singular thing comes into being and develops in perceptive experiences with the identical object of perception which I name Peter. It entails the usual elements of his appearance, like the shape of his face, his Gestalt, his voice and usual behavior, etc. Types come into being in concrete experience and they change constantly in further experience: If Peter used to have a black beard and suddenly he shows up being shaved, then my type of the singular thing Peter changes slowly to a person without beard.[7]

The type rests only on my experience of a limited number of perceptions, but it also has a kind of generality. In the case of a singular object like Peter, the resulting type entails different occasions, postures and different perspectives of him. Types change with my experience of the changing object. But the result of this process is not only a collection of incompatible views which can only show up in recollection and relatively rigid pictorial reminiscences. In the use of a type our former experiences are somehow dynamized: In types, our different experiences of the same object are connected in a kind of similarity-relation which allows us to identify this object as the same even as it appears in different ways to us. Thus in types we have not only a standard "front view" of a person, as in a mugshot of a criminal, but we have also the perceived experiences of one perspective view changing into another.

This dynamic similarity relation of perspective views and the transformations of these views should not be understood on the model of a tape recorder that passively collects information. Types allow us to transform, actively and independently from former experience, one appearance of the object into another with the help of imagination.[8] In my view, this transformation can happen in arbitrary imagination as well as in the form of a "helping phantasma".

In our arbitrary imagination, we are able to change the appearance of a person in one posture into another one and we are able to imagine the person laughing or talking, etc. But in the use of types in ordinary perception, we also make a very free and poetic use of the type in the sense of creating new and unknown perspectival

---

[7] We can realize that the type changes slowly by observe our slight astonishment when I see him the next time after he has shaved. He appears to us as something which should have a beard but has not. But the type slowly changes and in the end we expect him as beardless. This is only an example, you can use every singular object that changes his qualities or properties, like cars, pencils, the kitchen cooker etc. In observing carefully our attitude to the changing typical expectations we realize that the type changes only slowly in several equal experiences.

[8] There are some further aspects of what we call "similarity" which can be analyzed phenomenologically, cf. Lohmar (2004), 123–137.

views of the object. This use of the type is not arbitrary or dependent on my will. We can clearly realize this in the example of the person we "see" while having a phone conversation with him or the voice we "hear" while reading a letter from a friend. This ability reveals that our mind is not a "representational unit" which can only take exactly the same pictures out of the "archives" that are put into it in preceding experience. We are able to act quite freely with the material collected in experience and we can do so not only in arbitrary phantasy but also in the immediate use of perceptual types.

In this immediate use of types in perception we realize that our imagination is active in every perception. Think of the phone call with your friend. While you are talking on the phone with him you can "see" the person moving and gesturing with the help of phantasmata. In this he is perfectly moving according to the actual communication; moreover, his attitude and mood, revealed by the tone of his voice, is mirrored in his mimics and gestures that we "see" with the help of phantasmata. In this activity, our mind is projecting a completed image of what is perceived in this concrete situation: A well known person speaking in front of me, looking at me with all mimics, gestures and movements. All these activities are guided by the intuitive givenness in only one sensual (acoustic) field. But it also rests on our understanding of his speech and tone and our type of this person. Remember Kant's example of the face appearing in the curtain: Our phantasmata work as a perfect completion of the face. They fill-in exactly the missing traits to make the face appear completely. We can find the same function in the inner speech phantasmata while reading or thinking, even the sound of the voice changes with contents and mood of the author.[9]

If we generalize the performance of the helping phantasmata we realize that while perceiving an object we have phantasmata of all aspects of the object (or situation) at all times. This production of phantasmata is independent of the question whether the respective sensible aspects of the object are completely presented in most relevant aspects or not. Phantasmata are there in every perception but usually we do not register them, usually they are completely "overwritten" by sensuality. We will only register them when the corresponding field of sensuality is not filled appropriately.

But this results in the fact that we often have phantasma, even though we register them only in special situations, and this asks for an explanation. In my view the helping phantasmata are neither subliminal nor unconscious. I prefer to characterize as being "weak" because phantasmata are functioning in normal intentionality but are easily replaced or overwritten by real sensuality. This results in the fact that usually we are not aware of the phantasmatic elements of perception, but sometimes they can be registered.

But if it is true that phantasmata are there in normal perception, but they vanish immediately in concurrence with sensuality, then one needs to ask: Why must they vanish, why couldn't they stay in our perception? And why must they be there at all?

---

[9]Cf. Pessoa et al. (1998), 723–802.

Both questions have a systematic edge. The first I will try to answer with the help of evolutionary theory and the second with a short outline of the phenomenological theory of constitution with the help of types.

Phantasmata have to vanish, because from systematic considerations we must conclude that the helping phantasmata are not allowed to become stronger or even equal in intensity in comparison with sensual perception. Otherwise our expectation of what is to come by way of phantasmata will be concurrent with and perhaps even supervene sensuality. This will result in constantly erroneous or dubious perception. This should not be possible from an evolutionary point of consideration. Even a "vivid concurrence" of the helping phantasmata with the fulfilling sensual data does not make sense. It would destroy the security of our perception. But security of perception, right or wrong, as a basis for our acting is an evolutionary necessity. Therefore helping phantasmata must be distinctly weaker than sensuality.

But why must the helping phantasmata be there at all? This becomes clear with a view on perception guided by types: Usual perception is a complex synthetic performance which takes up sensual information from different fields and involves synthetic connecting acts. For example, when I see the person in front of me sitting behind a desk, I perceive a "complete" person, but in sensuality there is the desk in front of her and this desk cuts the "sensual givenness" of the person in two. I have to overcome this discontinuity in sensuality in my perception, therefore somehow I have to know whether a person can appear in this way or not. Thus I must have a kind of anticipating pre-image of a human person in a certain posture and in every perspectival view when I connect the sensual givenness actively in order to constitute a representation of the object. This activity of "creating a representation out of the sensual material by choosing, interpreting and combining" is often disregarded: The usual idea of perception is that every part of the object is passively given in sensuality and thus the object simply "appears" to us. Contrary to this simple way of appearance, a careful analysis of the conditions of the possibility of intuitive perception shows that we have to perform complex synthetic activities in every perception: We have to choose, interpret and connect in the material of sensuality so as to represent the object. We are not aware of this activities of the mind but we are able to reconstruct them in reflection. For example: Perceiving a human person, we cannot take the floor as part of the representing material, we cannot take our slight toothache nor the ring of a doorbell as a part of the representing material – but all this might be present in sensuality while we perceive the person. In connection with the choice of elements of sensuality, we also have to interpret these elements (e.g. a "line" or a smell) as a part of the presentation of the object seen.

After this *choosing* on the basis of our knowledge of how a person appears, we have to *combine* what we have found in different fields of sensuality to form a representation of what we are expecting to see. From the point of view of genetic phenomenology this guiding knowledge is hidden in the "type" we are using in the process of apperception. Think of the lemon appearing visually as a yellow area of a certain shape, but also in other dimensions of sensuality. We may have at the same time a faint fruity smell that we will connect with our perception of the lemon after *interpreting* it as a part of the usual appearance of a lemon sedimented in the type of a lemon.

In all these related activities of choosing, interpreting and combining, we must know what we are allowed to choose, how we have to interpret the parts and how to combine the sensual material, and for this we have to know how the lemon looks and smells in advance. We have to know what smell we are looking for – and in my view the way to know this is with the help of weak phantasmata of the smell of the lemon. If it happens that one field of sensuality is not filled, we might even be able to register the weak helping phantasma because of their necessary function in perception. Thus our analysis runs into the thesis that weak phantasmata play a decisive role in normal human perception: They always foreshadow an idea of a complete object perceived in all fields of sensuality.

If this is true, then what will happen if a complete field of sensuality is slowly decreasing in intensity up to the point where it completely faints? The conclusion is unavoidable that in this situation in normal people there will arise complex hallucinations. This consequence is confirmed by empirical psychology. The most frequently reported cases of complex hallucinations in psychologically normal people are to be found in persons who are slowly going blind, the so-called Charles–Bonnet–Syndrome (CBS).[10] The persons concerned are hallucinating predominantly in everyday situations other persons, animals and things, but sometimes also non-mundane things. They are nearly always well aware of the non-reality of the contents of their hallucinations (over 85%). Recent investigations shows that CBS arises reliably in 10–30% of all psychologically normal people slowly going blind.[11] But the percentage in reality may be even higher, because most of the CBS-patients were hesitant to report their hallucinations to doctors and other persons for they feared to be treated as mentally ill. Other investigations reveal that under comparable circumstances in normal people there are complex hallucinations possible also in other senses, the reports concentrate on verbal and musical hallucinations.[12]

On the basis of our analyses of the role of phantasmata in perception, we can propose an epistemological argument why such hallucinations must occur in senses that are weakened for a time or that fail completely. So we can argue why from this perspective CBS is "normal" and even makes sense. In our phenomenological analysis of perception I have pointed out that the human mind is able to insert helping phantasmata in all senses. This function is a necessary condition for the process of perception. It is a normal and useful function within perception, but usually it is unobtrusive: The phantasmatic products are usually overwritten by sensuality as long as the sense organs function normally. Phantasmatic self-affection is anticipating what is going to be present if we are oriented to the typical way in which a specific thing is going to present itself.

---

[10] Cf. Bonnet (1760). Some of the most recent systematic studies are: Teunisse et al. (1994, 1996); Podoll et al. (1989); Schwarz and Vaghei (1998); Schulz and Melzack (1991). Cf. also the interpretation of CBS in Ramachandran and Blakeslee (2002).

[11] Cf. Schulz and Melzack (1991), 813.

[12] Cf. Hosty (1994), 29, Griffith (2000), 2065–2076 and Berios (1991), 356–360.

Thus we have to realize that there is a complex two-sided interrelation or dependency of phantasmatic self-affection and intentionality. We might say that intentionality depends on phantasma as a representational mode of intentions towards objects and their qualities. And the contents of the helping phantasma are dependent on the kind of object intended. But we cannot identify both dependencies because the phantasma presenting parts of objects not actually perceived are dependent on type of object perceived. They fit perfectly in the other parts and form a *completion* of the sensual givenness of the object. But how is intentionality dependent on phantasma?

Intentionality describes the basic way our consciousness is directed to an object (or event). Intentions are fulfilled by given sensuality or they are empty intentions using signs or pictorial consciousness. There is a continuum of adumbrations between sensually fulfilled and empty intentions. But even in fulfilled intentions of an object there are still parts only emptily intended, like the backside or the interior. Especially in these cases, the emptily intended parts are not presented by the means of signs but in the way of weak phantasma. If we think of a carpet which is partly seen and partly under a cabinet, we might have the impression that the texture and the pattern of the carpet seems to continue under the cabinet.[13] Animate beings not able to use signs or words are intending unseen parts of perceived objects only with phantasmatic means.

Thus the whole presentation up to now reveals a theory of the principles of perception and knowledge that fits all animals that have the following three abilities: sensuality, the ability to reproduce sensual impressions in phantasmata and the ability to connect synthetically these elements so that they can constitute objects with properties.

Thus in my view most animals must be able to reproduce sensual impressions in the form of phantasma, like we humans do. But this suggestion is based on introspective description and transcendental analysis of the conscious life of the only animal we know from the subjective "view from within". In full generality this suggestion can only be proved in empirical psychological testing.

Nevertheless, there are arguments for this thesis: dogs and most mammals can dream. We notice their attempts to run in sleep, making noises and mimics of anger, fear and pleasure, etc. and we interpret this like our own dreams. Another argument from evolution theory: All animals have enemies (and preferred food) and thus all animals need the ability to recognize their enemy also under difficult circumstances or in an adumbrated situation.[14] An animal living in an Asian jungle must be able to recognize the tiger from very small hints, it must be able to "see" the tiger in a small moving spot of light, for it cannot wait to run away until the full shape of the tiger is in sight. In humans weak phantasma are also at work in situations when we fear a certain event: When a man enters a hut where a rope is lying rolled up in a dark corner, the man runs out of the hut because he has "seen" a snake.

---

[13] This example is used by Husserl in his *Logical Investigations* (Hua XIX, B 40).

[14] The same argument holds for preferred food and sites.

Most philosophical theories of knowledge presuppose that language and concepts are the basis for gaining knowledge. Our analysis of the function of weak phantasma in the human mind shows a viable alternative to this conviction. My hypothesis is that in most animals the constitution of perceptual objects is similar as in humans, i.e. it employs types mediated with weak phantasma. In the following I will show that this also opens up a new way to understand the remarkable continuity between primate's and human's principle cognitive abilities beside the equally remarkable and lasting difference in their factual performances.

Phantasmatic, Non-linguistic Modes of Thinking in Humans and Animals

The structure of my argument follows four theses which I will briefly present and then defend sequentially: (1) Language is one system of representation for cognitive contents, but we can in principle conceive of alternative systems of representation with the same performance. (2) There must necessarily be non-linguistic systems of representation in humans and in higher primates. (3) Non-linguistic systems of representation are still functioning in our own consciousness. We simultaneously use different systems of representation, among which are language, gestures, feelings and scenic phantasma. In regard to these theses, it is especially fruitful to investigate the scenic mode of daydreaming as a central form of non-linguistic mode of thinking. (4) As a consequence, it is highly probable that primates are able to think using the same non-linguistic systems of representation as we do.

In regard to the first thesis: At first sight human thinking seems to always use conceptual language. And there are some quite precise phenomenological descriptions of how we think with the help of concepts. Most basic in this regard is the insight into the cooperation of two kinds of acts. The first kind forms the ground for the intuitive evidence of states of affairs, which Husserl names categorial acts. The meaning-bestowing (*sinngebende*) acts are meant to connect this intuition with elements of a representational system like language. In this complex interplay of meaning-bestowing and intuitive, meaning-fulfilling acts giving the evidences of categorial objects, the most important movement is that of adjusting the expression to the intuition and not the other way around.[15]

Nevertheless, we are able to interpret expressed language as words and sentences that point to the intuition usually connected with these sentences. So we can find the corresponding intuition of cognitive objects at which words and judgments aim. But this also shows that language and the intuition of states of affairs, besides the fact that they are usually closely connected, are nevertheless not inseparable. Language is a certain system of representation of states of affairs originally intuited.

---

[15] For Husserl's theory of meaning cf. I. and VI. *Logical Investigation*. For the theory of categorial intuition cf. Ch. 6. of the VI. *Logical Investigation* and Lohmar (2002), 125–145.

But in contrast to linguistic representation, this intuition itself is more basic, originary and independent. With the help of language we are able to conceive the same states of affairs that we have had intuitively before, and this is possible even in the absence of intuition. This is, generally speaking, the central function of a system of representation.

But as we realize the difference between spoken words and intuitions of the states of affairs, sometimes we also realize the difficulty of adjusting linguistic judgments to the intuitive evidence (and also vice versa, of judgments understood to their corresponding intuition). This gives us a clear hint that language is only one of several possible systems of representation for cognitive contents operative within our thinking. Such a view finds support in Husserl's theory of meaning and in his theory of intuiting states of affairs. Categorial intuition is already an intuitive intention of a cognitive object, a state of affairs, a relation, a consequence, etc. The ability to transform this intuitive intention into an empty intention of the same object is the concept of "thinking" I will use. As we already see, thinking does not necessary have to be linguistic.

In Husserl's theory of meaning, there is still a tension between the categorial intuition and the appropriate linguistic expression that is best conceived of as striving towards an ideal norm of being the most appropriate true expression.[16] The concept of rightness names the aim of this striving, and the movement of approximation of the most appropriate expression is guided by the categorial intuition as a primary, self-standing and autonomous ruler.[17] This norm in intuition is also the leading tool in the use of other than linguistic means for thinking. "Thinking" denotes the principal ability to intend emptily (and manipulate) the object of cognition fully given in categorial intuition. Thus a "system of representation" must only allow for this function of thinking in a singular subject. If this system of representation allows also for public communication, like gestures and language, then it makes possible the conception of the same cognitive object also in other subjects. This is not the case for all systems of representation of cognitive contents, as we will see in the case of scenic phantasma that only support singular thinking.

Now we see that the inner or public expression of intuited categorial objects can use different means of which I will name three basic types with their performance and their characteristic limitations: Language and codified gesture languages (e.g. ASL, …), useful for singular thinking and public communication; non-codified gestures together with mimics, onomatopoeisis and pantomime, useful for singular thinking and public communication; scenic phantasma of past and future events are suitable for representation in singular thinking but not for public communication. Such scenic phantasma are not only to be found in our nightly dreams but also in our daydreams.

We see that the connection of language and thinking is not as narrow and firm as we tend to believe. Not only can we express our insights in different languages,

---

[16] Cf. Hua XIX/1, 313.

[17] Cf. Hua XVII, §46.

but we can also think in a language that is not our mother tongue. Most of us have experienced that after some time in a foreign country whose spoken language is one we well know, our thinking switches to the other language. It thereby becomes clear that the level of language is very close to the surface of the whole phenomenon of thinking and expression. The most basic level is categorial intuition.

But why do we have to use a symbolic medium at all? I think that this is due to the fact that we can hold on to the intuition of states of affairs only for a short time. After this we must use a symbolic medium to hold on to the meaning of our cognition. At the same time the intuition transforms into a firm conviction – which obtains also a symbolic form – that this state of affair is the case. This is even more true for the hypothetical manipulation of future states of affairs which we embark upon while thinking through our options.

Thus the symbolic carrier of a conviction is the presupposition for three essential performances of thinking: (1) the ability to awaken and to retain in mind the same object of cognition; (2) the ability to engender other cognitions from this one; (3) the ability to manipulate our future possibilities (and also different hypotheses concerning the course of history in the past). These central performances allow me to manipulate the possible future of an object or event in different situations, ponder possible consequences, obstacles and solutions of problems. Essentially, thinking is an active treatment of the contents of our cognition.

If we understand thinking as the ability to awaken, hold on to and manipulate the contents of cognition even in the absence of the intuition of the states of affairs, then we cannot deny that thinking must have a medium of symbolic representation. The latter, however, need not be language. Yet language gives us a clue of the most important features of such a system of symbolic representation: I must be able to produce the material carriers of symbols at any time; for example, I must be able to produce spoken or written words at any time either in public speech or in inner speech. I am able to think only if the symbolic carrier is ready at hand all the time. Following the pattern of Husserl's theory of meaning, this linguistic or non-linguistic carrier must achieve its meaning in a meaning-bestowing act based on the intuitive cognition.

Thus we may conclude what we already know: language is a usable carrier of cognitive meaning, it makes thinking and public communication possible because I can speak aloud any time. And in regard to inner thinking, I can let my inner voice function as the carrier of singular thinking. But our conclusions can go also beyond this trivial insight. The general feature of symbolic systems useful for thinking is only that I must be able to produce the carrier of symbols at any time – in inner or outer sensibility. Thus there can occur also internal carriers of meaning that allow for thinking but do not allow for public communication. Moreover, there may be also carriers of symbols that allow for both, such as language, gestures and pantomime. But it is obvious that language need not be the carrier in all these cases; there are always alternatives.

Now let us turn to the second thesis: There must necessarily be non-verbal systems of representation in humans and in higher primates. Two arguments support this thesis. The first argument stems form the history of human evolution.

To understand the second argument, we will have to consider modern research in cognitive abilities of primates.

The first argument centers on the problem of human evolution. A gap in our understanding of human evolution stems from the insight that spoken language and concepts cannot be older that 120,000–150,000 years. The decisive finding is that up to now there are no older findings of the tongue-bone of *homo sapiens sapiens*, i.e. the bone that enables refined phonetic languages and cannot be found in primates and older anthropoid species living before modern humans.

On the other hand, we know from the analysis of the lifestyle of early hominids that there must have been very powerful mental means for planning and also communicative means for the organization of cooperation already 2.5–1.8 million years ago. The basis for this conclusion is the knowledge that *homo erectus* settles the whole world at this time. Settling regions like deserts, tundra or northern Europe presupposes foresighted thinking, disciplined and flexible cooperation, social institutions and an extensive tradition of expert knowledge. All this has been possible without the use of spoken language.

We cannot avoid the conclusion that there must have been non-verbal systems of representation used for thinking and also for public communication. An influential line of thought in modern evolutionary theory suggests that it was gesture language that allowed public communication in these early hominids.[18]

Now the question arises whether these non-verbal modes of thinking and communication are still functioning in humans today, or whether they have simply vanished with the emergence of spoken language. If they still function in human consciousness – and this is my view – then we should be able to reveal this by means of phenomenological investigations. In my view, this non-verbal system of representation of cognitive contents operates with scenic phantasma and feelings (and beside this also with co-feelings representing sensations, feelings and intentions of other persons).[19]

But how can we conceive of the interplay of different systems of representation in one and the same subject? The thesis that in human consciousness there are two different but closely related processes which have the same mental performance has been put forth already in 1975 by P. C. Wason and J. St. B. T. Evans.[20] The first process is a *low level-system*, that is *phylogenetically old, relatively simply structured* and *less trouble-prone*. It enables quick insights and self-assured acting. This *low-level-system* is not language based and we have it in common with most of higher organized animals, whereas the *high-level*-system rests on the use of language and concepts. Therefore the *high-level*-system is *relatively slow* because it makes use of *complex rules* and *dependencies*. With the help of this conceptual

---

[18] Cf. Corballis (1999), 138–145. This thesis was proposed already by Condillac and Gorden W. Hewes.

[19] Cf. on the aspect of co-feeling and our way of understanding others, Lohmar (2006), 5–16.

[20] Cf. The contributions of Wason (1975), 141–154; Evans (1982), Ch. 12; Evans (2003), 454–459. A good overview about the different dual-process theories offers Stanovich and West (2000), 645–726, Ch. 6 ("Dual Process Theories and Alternative Task Construals.").

system we are able to constitute new abstract objects and higher level concepts. But most of our cognitive, emotional and volition performances we are also able to perform in the *low-level-system* which we share with many higher animals.

Let me now turn to the argument for non-linguistic systems of representation on the basis of the remarkable mental abilities of primates. That primates have knowledge about properties of objects and the probable causal results of events is already entailed in the use of tools. There are traditions of using and producing tools which sometimes endure for long times (the tradition of cracking nuts is 4,300 years old).[21] Beside this, it turns out that chimpanzees and bonobos are able to understand and to use languages based on abstract symbols and also on codified gestures. With both methods, they were able to form correct sentences of two to three words (which is the level of 2 year old children).[22]There is evidence that primates are able to have insight into the intentions and sensations of other mind's. They have an idea of their own (changing) social role in the hierarchy and of the "image" that others have of their knowledge and intentions. This becomes quite obvious when faced by the impressive documentation of deception in primates.[23] Beside this, there are also rudimentary moral rules of different kinds in the complex social systems higher primates are living in.[24] They are individuals with personal history and character. They have an idea of their outward appearance and the conclusions that other members of their group will draw from their outer appearance, mimics and gestures. Moreover, they are able to intentionally influence their outer appearance to manipulate the opinions and the behavior of others (for example in deception by false limping).[25] Primates are capable of many kinds of technical, social and political cooperation.

As we now have a clear idea of non-linguistic thinking, we may also wonder how primates may think. Our claims regarding this issue will be based on the phenomenological analysis of our own non-linguistic thinking. Thus through these analyses we might perhaps also discover in which way we are still thinking like animals.

We used to belief that humans perform all (or at least the most important) of their mental performances with the help of language. But our considerations rendered this belief questionable. It may be that simple systems of representation can engender similar results and that we share these basic non-linguistic systems of representation with primates. Therefore public language may be only a supreme layer of representation not really contributing to our thinking. I will not deny that language allows us to refer to abstract things and cognition on a very general level. Regarding the formal aspects, for example our claims for necessity and universality, language allows us to speak about relations that we can only address with the help

---

[21] Cf. Mercader et al. (2007), 3043–3048.

[22] Cf. Fouts (1997) and Savage Rumbaugh and Lewin (1994).

[23] Cf. Whiten and Bryne (1986), Whiten and Bryne (1988a), Whiten (1988b) and Sommer (1992).

[24] Cf. Frans de Waal's writings on the social life of primates, de Waal (1982, 1989, 1996).

[25] This is reported by Frans de Waal, cf. Bryne and Whiten (1990), 1–101, episode 238.

of language. But it might turn out that by speaking on this general level we can only speak about cognition that we have gained with the help of simple low-level systems of representation.

Nevertheless, there is a factual difference between the cultural performances of primates and humans. This is due to the possibility of public communication opened up by language. Language enables an immense advance for factual human performances: communication allows us to transfer the insights of single members of the group to all others and it allows us to accumulate technical and social inventions so as to solve problems with the help of tradition, which remains an abiding possession from one generation to the next. Even if we have to realize that other primates have a comparably inventive intelligence, the human communicative and traditional facilities prevent us from forgetting what we already know about useful technical and social tools. Here we face an obviously important factual difference. M. Tomasello has pointed out the important performance of this "ratchet-effect" through tradition in human social and technical culture.[26]

Let us turn to the third thesis: Non-linguistic systems of representation are still functioning in our own consciousness. We simultaneously use different systems of representation, among which are language and codified gestures, non-codified gestures, pantomime, emotions and scenic phantasma.

Our ability to communicate by means of non-codified gestures and pantomime is broadly underestimated. Consider having to go to the airport in a foreign country whose language you cannot speak. Consider, further, an encounter with a taxi driver whom you need to inform about your urgent wish. In a situation like this one, one immediately starts communicating with the help of gestures, onomatopoetic means and pantomime.

This behavior is very informative of our non-verbal systems of representation. We start without further thinking and we are certain about our attempt to communicate in this way. And this reveals that this non-verbal mode of communication is still alive all the time while we use language, for we do not have to wonder about the "how" of this gestural-pantomimic communication. We simply start using it, as if we have been tacitly doing it all the time. This, however, is only an example for public communication without spoken language. Now I would like to turn to non-linguistic modes of internal thinking.

We use scenic phantasma as expressions of our wishes and fears in our daydreams and thus they function as representations of cognitive contents. It may be a state of affairs that we wish for or are in fear of. But we do not thereby simply express our preferences and our views of the state of affairs; daydreaming is also a kind of mental action in regard to addressing problems.[27]

In daydreams we play out possible solutions to a problem by way of mentally testing our options, their usefulness and their respective consequences. This life of

---

[26] Cf. Tomasello and Call (1997), Tomasello (1999).

[27] We might object that our daydreams are completely free and arbitrary, but this is not the case as is shown by endless repetitions of the same motives in daydreams expressing fears. Cf. Lohmar (2008), Kap. 9.

scenic phantasma constitutes an important, yet broadly underestimated part of our conscious life: Worries about urgent challenges or uncertainties that make us sleep-less at night. There are many phantasies of having success. I would like to mention also empirical-psychological research that suggests that most grown up males think of sex every few minutes, and the mode of this thinking is definitely not concep-tual.[28] In these scenic episodes of our conscious life, the linguistic expressions recede into the background in favor of pictorial elements.[29] I am not denying that we can also think about our wishes and problems with the help of language and that in daydreams both are often merged, but what I want to stress is that we also use non-linguistic systems of representation.

Most higher-developed mammals can dream. They show first signs of an attempt to act and emotions in the phases of their sleep, which we interpret as episodes in dreaming that prolong wakeful states of goals and action. We might therefore claim that the representation system with scenic phantasma is operative in primates in dreams and wakeful states in the same way as in humans.

We might also interpret our feelings as an important element of non-linguistic systems of representations, functioning in the framework of scenic phantasma. Perhaps we cannot interpret emotions as a self-standing system of representation, because we must always presuppose intentions of events to which feelings are sub-sequently directed. Emotions can easily be granted the status of the most important system of non-verbal representation for we can have them in an actual situation and we can "produce" them (not arbitrary) also in the absence of the intuitive situation only by imagining it. The feeling of fury is moving me violently in a certain situa-tion but also in mere thinking of the same situation later on. In both cases, the feel-ing "tells" me something about the value of the event, it is a part of my inner "expression" that has a certain meaning. In thinking about a pleasant experience the agreeable feeling "means" the desirable quality of the event.

But now let us turn back to daydreams that also perform a representation of our wishes and fears. They mirror our personal order of significance between the two poles of events that should never happen and that should happen at any costs. They do not ask for a refined psychoanalytical hermeneutics (at least at first glance). Contrary to nightly dreams, daydreams respect the identity of objects, causality and order in time. Therefore, daydreams can be interpreted as a thinking activity of the present and future reality.

The framework of our order of relevance in possible events makes us also under-stand better why special daydreams must be experienced for as long as the urgent needs and fears remain the same and unaltered. But we have to be attentive to small modifications in these repetitions that represent my possible options in real action.

For example: Had I been pressed hard by an impertinent and aggressive guy and had I given way to his demands due to the situation and circumstances, this annoying situation would reemerge repeatedly in my furious daydreams. This reappearance

[28] Cf. Cameron and Biber (1973), 144–147.
[29] This is also true for nightly dreams, cf. Symons (1993), 181–217.

would also engender the right solution to do away with his aggressive demands: "This would have been the right reaction; had I done this, it would have stopped him!" This insight is, however, irreal and it cannot change the past. And yet, it is a kind of engagement with reality that enables me in a future situation to act appropriately. The same is true for events I am anxiously expecting.

Thus the special scenic mode of daydreams allows for an interpretation of daydreaming as an old mode of thinking. If I am worried in the mode of daydreaming then things and persons are occurring in pictorial representation and language shifts into the background. The content of my worries is represented in scenic phantasma, but every time with small modifications. And in these modifications we sometimes realize successful solutions to our problems: Winning in a lottery will solve some problems; working hard or suffering for some time from some privations will bring the right results. This shows clearly that the function of daydreaming is that of a non-linguistic mode of thinking that can somehow "move" all problems in thoughts towards their possible solution. I will not deny that in turning back from our inner life of scenic phantasma to other members of our group we will immediately change to a language-mode of communication, but this shift will only express what was already found with scenic means before.

Let me turn to the fourth thesis. There is only a limited set of themes that primates living in groups have to be able to think about. (1) Objects, their present and future states and use (e.g. as a tool), as well as their value in my personal estimation and their value in the view of the community, i.e. the cultural value. (2) Events in the present, past and future, their felt value and their probable consequences. (3) Other persons with their sensing, feelings, convictions and their practical intentions related to me and other members of the group. I hope that I can leave it to the reader to find scenic phantasmata representing the first two themes and concentrate on the last group of intentions of other persons.[30]

It seems difficult to imagine a scenic phantasma of the character of a person and of his or her probable behavior towards me, especially within complex constellations with others who are involved in the action. But scenic phantasma offer a simple solution for this apparent difficulty. In remembering a brutal former classmate, I see his face looking at me with evil eyes, with clenched fists, and ready to give me a beating. But this "image" is not simply an image of him, it is a *characteristic scene* within which I am present, writhing with pain from his beating and in fear of further beatings. This scene presents central aspects both of his character and of his future behavior.

But scenic presentation of the attitude and the behavior of a person need not be so one-dimensional as in the case mentioned, since normally there are multiple facets in the character of other persons. But how can I think a multitude of (changing) attitudes in a scenic mode? Think of a colleague with whom you work together successfully in most cases, but who occasionally appears with an air of high-nose arrogance. Both "faces", i.e. both aspects of his character, may be represented in a

---

[30] Cf. on co-feeling and our way of understanding others Lohmar (2006), 5–16.

scenic phantasma, one after the other, or even, as mixed in a changing way, which results in an uncertain base for your plan-making. The character of possibility and uncertainty is thus present in the changing and merging faces of your colleague. We might interpret this changing image as a non-linguistic form of the logical „or".[31] His attitude towards other persons and other situations may be represented in a similar way since you can easily extend the characteristic scenes.

The value and the use of objects can change, which is also reflected in the characteristic scene. For instance, if I own a car that usually breaks down and thus has to be towed and repaired, the characteristic scene within which I am positively excited about my car is modified, and converted to one that is negative. The emotional aspects of this bad experience are especially mirrored in the characteristic scene: I no longer imagine the car with the joyful expectation of reliable use, but with the cheerless expectation of future harm and inconvenience.

To conclude: Scenic phantasma in daydreams constitutes a central non-language system of representation, an "old mode of thinking" employed by humans and probably also by primates. Our analyses of non-linguistic modes or thinking delimits the significance of language for human thinking in clear way. Language is by far not the only possible means of thinking and, moreover, it is not the only system of representation operative in the human consciousness. It seems to be probable that the real basic performances of cognition and our conception of reality is based on more simple phantasmatic systems of representation that are still operative in our mind. Public language and the concepts it uses are only a very superficial layer of the whole performance of thinking. Thereby a new task for phenomenological research is opened: to analyze the modes of alternative systems of representation in human consciousness.

References

Berios GE (1991) Musical hallucinations. A statistical analysis of 46 cases. Psychopathology 24:356–360

Berthoz A (2002) The brains sense of movement. Harvard University Press, Cambridge, MA

Bonnet C (1760) Essai analytique sur les facultes de l'âme. Kopenhagen, C, Philibert A (eds) 2nd edn. Philibert, Copenhagen

Bryne R, Whiten A (1990) Tactical deception in primates: the 1990 database. Primate Report 27

Cameron P, Biber H (1973) Sexual thought throughout the life span. Gerontologist 13:144–147

Corballis MC (1999) The gestural Origins of Language. Am Scientist 87(2):138–145

de Waal F (1982) Chimpanzee politics. Jonathan Cape, London

de Waal F (1989) Peacemaking among primates. Harvard University Press, Cambridge, MA

de Waal F (1996) Good natured. The origin of right and wrong in humans and other animals. Harvard University Press, Cambridge, MA

Evans J St. BT (1982) The psychology of deductive reasoning. Routledge and Kegan Paul, London

---

[31] Logical operators belong to the second class of objects that primates must be able to think about: Present and future events, with their value for us together with their probable consequences.

Evans J St. BT (2003) In two minds: dual-process accounts of reasoning. Trend Cogn Sci 7:454–459

Fouts R (1997) Next of kin. What chimpanzees have taught me about who we are. William Morrow, New York

Griffith TD (2000) Musical hallucinosis in acquired deafness. Brain 123:2065–2076

Hosty G (1994) Musical hallucinosis: Auditory Charles Bonnet syndrome? Ir J Psychol Med 11

Kant I (1764) Versuch über die Krankheiten des Kopfes. In: Kants Werke. Akademie – Textausgabe. Bd. II. de Gruyter, Berlin, pp 1923

Kant I (1766) Träume eines Geistersehers ergänzt durch die Träume der Metaphysik. In: Kants Werke. Akademie – Textausgabe. Bd. II. de Gruyter, Berlin, pp 1923

Lohmar D (2002) Husserl's Concept of Categorical Intuition. In: Zahavi D, Stjernfelt F (eds) Hundred years of phenomenology. Kluwer, Dordrecht, pp 125–145

Lohmar D (2004) Wittgenstein, Husserl and Kant on aspect-change. In: Benoist J, Laugier S (eds) Husserl et Wittgenstein, Olms Verlag, Hildesheim

Lohmar D (2006) Mirror neurons and the phenomenology of intersubjectivity. Phenomenol Cogn Sci 5:5–16

Lohmar D (2008) Phänomenologie der schwachen Phantasie. Untersuchungen der Psychologie, cognitive science, Neurologie und Phänomenologie zur Funktion der Phantasie in der Wahrnehmung

Mercader J, Barton H, Gillespie J, Harris J, Kuhn S, Tyler R, Boesch C (2007) 4300-year-old chimpanzee sites and the origins of percussive stone technology. PNAS 104(9) February:3043–3048

Pessoa L, Thompson E, Noê A (1998) Finding out about filling-in A guide to perceptual completion for visual science and the philosophy of perception. Behav Brain Sci 21:723–802

Podoll K, Osterheider M, Noth J (1989) Das Charles Bonnet Syndrom. Fortschritte der Neurologie und Psychiatrie 57:43–60

Ramachandran R, Blakeslee S (2002) Die blinde Frau, die sehen kann. Rätselhafte Phänomene unseres Bewusstseins. Rowohlt, Reinbek, pp 177–190

Savage-Rumbaugh S, Lewin R (1994) Kanzi. The ape at the brink of the human mind. Wiley, New York

Schulz G, Melzack R (1991) The Charles Bonnet Syndrome. Phantom visual images. Perception 20:809–825

Schwarz TL, Vaghei L (1998) Charles Bonnet syndrome in children. J Am Assoc Pediatr Ophthamol Strabismus (JAAPOS) 2(5):310–313

Sommer V (1992) Lob der Lüge. Täuschung und Selbstbetrug bei Tier und Mensch. dtv, München

Stanovich KE, West RF (2000): Individual differences in reasoning: implications for the rationality debate? Behav Brain Sci 22(5)

Symons D (1993) The stuff that dreams aren't made of: Why the wake-state and dream-state sonsory experiences differ. Cognition 47:181–217

Teunisse RJ, Zitman FG, Raes DCM (1994) Clinical Evaluations of 14 Patients with the Charles Bonnet Syndrome (Isolated Visual hallucinations). Compr Psychiatry 35:70–75

Teunisse RJ, Cruysberg JR et al. (1996) Visual Hallucinations in psychologically normal people: Charles Bonnet's syndrome. CBS. The Lancet 347 March:794–797

Tomasello M (1999) The cultural origins of human cognition. Harvard University Press, Cambridge MA/London

Tomasello M, Call J (1997) Primate cognition. Oxford University Press, Oxford

Wason PC, Evans J St. BT (1975) Dual Processes in Reasoning? Cognition 3:141–154

Whiten A (ed) (1988b) Macciavellian intelligence, Oxford University Press, Oxford

Whiten A, Bryne R (1986) The St. Andrew Catalogue of tactical deceptions in primates. St. Andrews Psychol Rep 10:1–47

Whiten A, Bryne R (1988a) Tactical deceptions in primates. Behav Brain Sci 11:233–273

Part III
Embodiment

Myself with No Body? Body, Bodily-Consciousness and Self-consciousness

Dorothée Legrand

De facto, subjects are bodily, but is this necessarily so? This question unfolds into the following two: Can one be self-conscious without being a body? Can one be self-conscious without being bodily-conscious? In this paper, I will describe different aspects of bodily-self-consciousness and argue that the most radical attempts to demonstrate the contingency of the bodily nature of self-consciousness failed. The idea here is that the strength of the anchoring of self-consciousness in the body and bodily-consciousness can be evaluated by considering whether (and which) forms of bodily-self-consciousness resist even against radical theoretical and clinical cases of purported disembodiment.

A Certain Unity

Cartesian dualism intends to radically eliminate the body as Descartes argues that "my mind, by which I am what I am, is entirely and truly distinct from my body, and may exist without it" (1641, VI, 9). Importantly, he also insists that "I am not only lodged in my body as a pilot in a vessel.… I am besides so intimately conjoined, and as it were intermixed with it, that my mind and body compose a certain unity" (1641, VI, 13).

One can see that these two points are not in blatant contradiction with each other, by distinguishing the following questions: (1) "Is the mind entirely and truly distinct from the body?" and (2) "Am I entirely and truly distinct from my body?" Question (1) concerns *the* mind and *the* body considered abstractly-generically, and Descartes argues that they are conceivable in isolation from each other; that's the core of his dualistic conception of the *Res Cogitans*.

D. Legrand
Danish National Research Foundation: Center for Subjectivity Research,
University of Copenhagen, Denmark;
Chercheur CNRS
CREA: Centre de Recherche en Epistémologie Appliquée,
ENSTA-32, boulevard Victor
75015 Paris, France

In contrast, question (2) concerns the mind and body of a concrete-specific subject and Descartes himself argues that the latter is a mind–body composite. Importantly "body as part of the union should not be confused with body as the extended substance; and the same holds for the soul: as part of the union it is not merely the principle of thinking" (Heinämaa 2003, 28). A factual subject's body is determined by its union with his mind: despite quantitative physical changes, it remains a subject's body throughout his life.[1] Likewise, *the* mind can be abstractly conceived as a purely spiritual substance, but a factual subject's mind is nonetheless linked to his body "substantially,"[2] and not only contingently. Therefore, we can conceive of a man without body only if we tackle the "metaphysical mind". If by contrast, we tackle the factual subject, "a handless or sexless man is as inconceivable as one without the power of thought" (Merleau-Ponty 1945, 197).

Despite his radical dualism, Descartes himself thus fails to disembody the mind of the factual subject and his view rather calls for further investigation of the bodily-self. In what follows, I will consider the factual subject and "the necessity of [his] concrete and contingent existence in the midst of the world" (Sartre 1943, 359). That I will consider the factual man does not imply that I will restrict my investigation to descriptions of factual links between body, bodily-consciousness and self-consciousness. I will rather discuss the necessity of such factual links by addressing the two aforementioned questions: Can one be self-conscious without being a body? Can one be self-conscious without being bodily-conscious?

Four Irreducible Bodily Dimensions

Not only dualism, but also materialism promotes a disembodied view of the mind. A radical version of this view is well-illustrated by the so-called "brain-in-a-vat" thought-experiment (Putnam 1982),[3] where a neurosurgeon separates your brain from the rest of your body. Your brain is maintained alive by being placed in a vat full of the necessary nutriments. It also remains related to the external world thanks to its connection to a computer. This brain-vat-computer set would be sufficient to generate full conscious experience. For example, the computer can activate relevant neurons in such a way that the person-brain (illusorily) experiences herself reading

---

[1] See, e.g. Descartes' Letter to Mesland.

[2] See, e.g. Descartes's reply to Arnauld.

[3] This thought-experiment was first developed to illustrate the sceptical problem of the existence of the external world.

these words aloud.[4] As illustrated by this "intuition pump", neuronal reductionism intends to isolate the physical correlates of self-consciousness in a subset of neuronal activations, thereby defending a "skullist" conception of self-consciousness (and/or its correlates) as in-brained but disembodied.

However, this disembodiment appears as far less radical if one considers this thought-experiment more in detail. First, the "thought-experimenter" would insist that the aforementioned surgery would disembody the brain of his "patient" but would not alter his subjective experiences. This notably implies that the patient would experience his body after the surgery (while he is putatively disembodied) just like he did before the surgery (while he was still biologically embodied).[5] If her experience is indistinguishable from a veridical experience, as the thought-experimenter wants it, the patient experiences herself as physically anchored in her body and world. Thus, the thought-experiment does not eliminate the experiential dimension of the body, characterized by the subjective experience of the body from a first-person perspective. This dimension comprises all aspects of what it feels like to be a body.[6]

Secondly, this thought-experiment does not concern a brain, but a brain-*in-a-vat*. The vat allows the brain's life-regulation, and for that to be effectively the case, such "vat" would have not only to contain the brain but also to simulate physiological life-regulating processes. The tentative disembodiment would thus fail to eliminate the anatomical dimension of the body, which comprises what characterizes and what goes on within the limits of the skin boundary.

Thirdly, this thought-experiment is not about a brain-in-a-vat but about a brain-in-a-vat-*connected-to-a-computer*. The computer allows the coupling of brain activations with the external world, and for that to be effectively the case, such "computer" would have to simulate the sensorimotor dimension of the body.

Fourth, in the thought-experiment, the body is also present at the neuronal level. Beyond the obvious claim that the brain is part of the body, the point here is also to underline that the "body" is in the brain (Berlucchi and Aglioti 1997), in the sense

---

[4]A classical interpretation of this thought-experiment can be summarized by the following voice: "[A] disembodied but appropriately stimulated brain in a vat could – phenomenologically – enjoy exactly the same kind of conscious experience as you do right now while reading [these words]. In principle, it would even suffice to properly activate just a subset of this brain, the minimally sufficient neural correlate of your present state, to make a 'phenomenological snapshot' of exactly the same kind of conscious experience emerge." (Metzinger, 2003, 547). Compare with: "I remark, in the next place, that the mind does not immediately receive the impression from all the parts of the body, but only from the brain, or perhaps even from one small part of it…" (Descartes, 1641, VI, 20).

[5]Putnam (1982) mentions that "the computer is so clever that if the person tries to raise his hand, the feedback from the computer will cause him to 'see' and 'feel' the hand being raised". Interestingly for the point at stake here, he adds a little later: "Of course, if we were two lovers making love, rather than just two people carrying on a conversation, then the suggestion that it was just two brains in a vat might be disturbing".

[6]See below (sections on The body-as-object and The body-as-subject).

that states and representations of the three aforementioned bodily dimensions figure among the conditions of activation of the brain. This defines the neuronal dimension of the body.

Note that this description is not meant to list exhaustively all possible bodily dimensions and that distinguishing these four bodily dimensions is not only relevant for better understanding an infamous thought-experiment. Rather, and more interestingly, these dimensions are dissociable from each other in real-life cases as well. For example, some cases of amputation are best characterized by a dissociation between the anatomical bodily dimension (i.e. absent limb) and the experiential (i.e. sensory ghost or phantom limb) and sensorimotor (i.e. the amputee moves and the posture adjusts as if the arm is still there) bodily dimensions.[7] Anosognosia for hemiplegia dissociates the sensorimotor (i.e. the hand cannot move) from the experiential (i.e. the patient reports he is moving) bodily dimensions. During tool use**,** the anatomical bodily dimension (i.e. the hand) is dissociated from the sensorimotor (i.e. the tool) and neuronal bodily dimensions (i.e. neurons activation previously correlated with the hand would now correlate with the use of the tool[8]), while the experiential dimension would be modulated accordingly.[9]

Such dissociations suggest that these four bodily dimensions cannot be reduced to each other. In the framework of cognitive neurosciences, this implies that the body and its actions are not reducible to their representation in the brain; one rather needs to also consider in their specificity anatomical and sensorimotor processes as well as the experiential dimension of the body.[10]

Importantly, these four bodily dimensions are not only irreducible to each other, but also irreducible *tout court*: they cannot be eliminated. Their biological implementation could not be replaced by just *any* artificial device but it could be replaced by artificial devices able to play a role *equivalent* to the role usually played by the biological body. These dimensions would thus be bodily, at least in a functional sense. Moreover, recent considerations show that, to be efficient at all, these artificial devices would have to meet important requirements by which they would in fact be surrogates of the biological body (Cosmelli and Thomspon 2009). Therefore again, it is justified to consider these processes as bodily.[11]

---

[7] See Brugger (2006).

[8] See Iriki et al. (1996).

[9] "Once the stick has become a familiar instrument, the world of feelable things recedes and now begins, not at the outer skin of the hand, but at the end of the stick" (Merleau-Ponty, 1945, 175–176).

[10] The irreducibility of the experiential dimension of the body to the other three dimensions is meant to be epistemological and does not imply any dualist view according to which the mind would be ontologically irreducible to matter.

[11] Cosmelli and Thompson (forthcoming) use the term "body" to mean a self-regulating system comprising its own internal, homeodynamic processes and capable of sensorimotor coupling with the outside world.

One important consequence of this interpretation of the brain-in-a-vat-which-turns-out-to-be-necessarily-embodied is that neuronal reductionism is at best incomplete, since the body (or any equivalent) is essential to the functioning of the brain itself. Even if abstract considerations allow one to conceive of the physical correlates of self-consciousness as located in the brain isolated from the rest of the body, concrete considerations of the conditions of realisibility of such correlations underline that the brain cannot be disembodied, and in turn, self-consciousness cannot be disembodied[12] (Gallagher 2005b). The brain-in-a-vat story is in fact self-refuting (since presupposing it true implies its falsity): Neuronal reductionism fails to disembody the brain since it requires a functional body (biological or artificial) at least as a condition of possibility for the functioning of the brain, i.e. according to its own hypothesis, for the emergence of subjective experience.

However, one may refrain from drawing the conclusion that not only self-consciousness cannot be disembodied but also, and as a consequence, the body should be considered when accounting for self-consciousness. Indeed, a neuronal reductionist could still maintain his "skullist" approach by arguing that the body is necessary for self-consciousness, but only secondarily: self-consciousness is not disembodied since it relies on the activation of the brain which itself cannot be disembodied, but what primarily matters for self-consciousness is cerebral activation. The fact that the body is necessary for the functioning of the brain would thus be irrelevant for the investigation of self-consciousness. Such neuronal reductionist would not apply transitive rules on body, brain and self-consciousness: he may agree that consideration of the body is necessary for understanding the conditions of activation of the brain, and argues that understanding the conditions of activation of the brain is necessary for understanding self-consciousness, but does not conclude that consideration of the body is necessary for understanding self-consciousness.

In reply, let me underline that this view misses the point of the previous discussion, as it wrongly reduces the body to only a remote condition of possibility for the brain to play its role in eliciting self-consciousness. This view simply neglects the fact that the body pre-processes and post-processes any signal getting in or out of the brain (Chiel and Beer 1997), in such a way that a given cerebral activation processed in or out of the body would not lead to the same outcome, in the behaviour and presumably the experience of the subject. In this sense, the brain is not self-sustaining but is a component of the system it forms by its intertwining with the body. Therefore, just as a single neuronal activation is not isolated from other cerebral activations, the brain as a whole isolated from its body couldn't play its functional role.

Moreover, it couldn't be properly understood either. Surely, one can correlate single-neuron activations to mental events, and isolating experience- and brain-slices is arguably the easiest way to do so. However, what does such slicing teach us?

---

[12] "The importance of the body can be measured in considering precisely what it would take to sustain a disembodied brain and the supposed experience that goes along with it" (Gallagher and Zahavi 2008, 131)

Knowing that brain and mind correlate with each other is a discovery only for dualists. Knowing how cerebral and experiential states correlate with each other is a more interesting issue, which necessitates understanding for themselves the terms of the correlation. On the one hand, understanding experiences requires considering their *structure*, in a phenomenological perspective which avoids their reduction to the *content* of introspective states[13] (Gallagher and Zahavi 2008). On the other hand, understanding cerebral states requires considering their intertwining with bodily states, intertwining by which the embodied brain forms the system constituting the putative "minimally sufficient" physiological correlates of subjective experiences (Cosmelli and Thompson, forthcoming; Thomspson 2007).

In reaction to this point, the neuronal reductionist may argue that "*as a matter of fact*", being anchored to one's body is not necessary for being self-conscious. To make this point, he would abandon the brain-in-a-vat thought-experiment, and turn to real-life cases. In particular, he may be tempted to exploit cases of Out-of-Body Experiences (OBEs). The latter occur in 5–10% of the general population and are defined "as brief subjective episode[s] in which the self is perceived as being outside the body (disembodiment), with or without the impression of seeing the body from an elevated and distanced visuo-spatial perspective (autoscopy)" (De Ridder et al. 2007[14]). Consider the following case-report: "Suddenly it was as if he saw himself in the bed in front of him. He felt as if he were at the other end of the room, as if he were floating in space below the ceiling in the corner facing the bed from where he could observe his own body in the bed. […] he saw his own completely immobile body in the bed; the eyes were closed" (Lunn 1970, case 1; in Blanke et al. 2008). OBEs are classically interpreted as cases where people "truly are disembodied thinking selves" (Metzinger 2003, 502), suggesting that "'I' can function without my physical body and see without eyes" (Blackmore 1982, 5).[15]

However, OBEs do *not* provide any argument for the neuronal reductionist's view according to which the body itself would not play any role in one's experience of oneself. Indeed, it is worth remembering what is too often overlooked: OBEs do *not* correspond to a dissociation of the subject from his biological body. OBEs are illusions, i.e. *experiences* of oneself as located out of one's biological body, and concern only the experiential dimension of the body. As underlined by Blackmore about her own OBEs "I was not functioning *without* my physical body. I seemed to be in a different place from that body, but there is no doubt that it was functioning quite well … it is therefore unjustified to conclude that such experience could take place without a functioning body" (1982, 5).

---

[13] One's body is taken as a *content* of intentional state when it is what I am conscious of (e.g. I look at my hand). Even when it is not the case, one's body *structures* one's intentional states both spatially and temporally, by defining an "here" and "now" to which one's perspective is anchored. See sections on The body-as-object and The body-as-subject for more details.

[14] See the glossary in De Ridder et al. 2007, 1830.

[15] "Such statements are common after spontaneous OBEs but they can be based more on emotion than reasoning" (Blackmore, 1982, 5).

Furthermore, it has been shown that direct electrical cortical stimulation[16] elicits OBEs if the subject looked straight ahead (without fixation of any specific object), the experience of limb shortening if she fixated her outstretched arms or legs, and illusory limb movement if the limbs were bent at the elbow or knee and illusory movement of her upper body towards her legs if she closed eyes (Blanke et al. 2002; Blanke and Thut 2006). Such empirical investigations of OBEs suggest that the anatomical body (its posture) does play an important role in eliciting various forms of bodily illusion, thereby defeating again neuronal reductionism.[17]

At this point, our objector may defend the conception of self-consciousness as at least potentially disembodied in another framework, where he would not argue in favor of a disembodied brain, but would rather directly argue in favor of a disembodied experience, e.g. by interpreting OBEs as non-bodily forms of self-consciousness. This leads us to the second of the two questions mentioned in the introduction: Can one be self-conscious without being bodily-conscious? Answering this question requires considering fine-grained descriptions of bodily-self-consciousness. That will be the task of the next two sections.

The Body-As-Object

As it is a *physical object*, one of the characteristics of the body is that it is opaque in the sense that one can look *at* it rather than *through* it (Legrand 2007e). Thereby, the body is taken *as-intentional-object*, i.e. as the object towards which one's intentional act of consciousness is directed. The body is taken as-intentional-object and is experienced as-physical-object, not only anytime I observe it through exteroception (e.g. when I look at it), but also when interoceptive signals call my attention and are thematized (e.g. when the body hurts) and when the body does not accommodate my projects and leads me to experience the resistance of the world (e.g. when I bump into some objects). These are all punctual but recurrent cases. Importantly, the body-as-intentional-object is unlike any other objects,[18] as it can be experienced as mine/me in a way other objects aren't.[19]

---

[16] Focal electrical stimulation at currents of 3.5 mA, for 2 seconds, of the junction of the right angular gyrus and the posterior superior temporal gyrus, in a 43-year-old right-handed woman, for intracranial presurgical epilepsy evaluation for intractable seizures.

[17] Contrast with: "a minimally sufficient neural correlate for the OBE state in human is likely to exist" (Metzinger, 2003, 503).

[18] See also Knoblich et al. (2006). See next section for a discussion of the idea that the body is not merely experienced as an intentional object, even an "extraordinary" one.

[19] In some cases, the body-as-object might be experienced but not recognized as mine. For example, a patient suffering from somatoparaphrenia may claim that his limb in fact belongs to another person. Such cases concern the possibility of failures of self-attribution of body parts, and is not the focus of the present investigation which rather concerns the inverse issue of the (im)possibility of being self-conscious without being bodily-conscious.

I also experience myself as-physical-object when my body is taken as-intentional-object by others. Sartre argues that "with the appearance of the Other's look I experience the revelation of my being-as-object" (Sartre 1943, 351). More precisely, the revelation would be to experience my body not only as-physical-object but also as-intentional-object-for-others, i.e. beyond my own intentional acts. As just listed above, the experience of my body-as-physical-object has itself different components, some of which can be revealed without others but rather in contact with the physical world. As my facticity is revealed by my experience of my body-as-physical-object, the limitation of my subjective perspective is revealed by my experience of my body-taken-as-intentional-object-by-others: because of others, "my body is there not only as the point of view which I am[20] but again as a point of view on which are actually brought to bear points of view which I could never take; my body escapes me on all sides" (*Id.*, 352).

Last but not least, to adequately account for the body-as-intentional-object, one must distinguish its appearance from its appear-ability. The way it appears (*what* appears) is contingent, since one's representation of one's body is always more or less distorted (compared to "objective" measures of the body). Nonetheless, the very fact that the body can be taken as-intentional-object by oneself and/or others should not be neglected nor conflated with the way it appears. In other terms, the fact *that* the body appears or can appear in consciousness is not reducible to *what* appears in bodily-consciousness.

The Body-As-Subject

The experience of the body-as-intentional-object is an "aberrant type of appearance" (Sartre 1943, 357) as it does not give us "the body as its acts and perceives but only as it is acted on and perceived" (*Id.*, 358). If that were the only form of bodily-consciousness, the subject would merely be *embodied* ("an I, to which the body belongs"; Husserl 1989), while he is *bodily* and experiences himself as such. This point can be detailed by applying Wittgenstein's distinction between I-as-object and I-as-subject (1958, 66–67) to bodily-self-consciousness (Legrand 2006). For example, when I look at my image reflected in a mirror, the *perceived self* corresponds to the body-as-object (*Körper*), while the *perceiving self* corresponds to the body-as-subject (*Leib*).

The distinction between body-as-intentional-object and body-as-subject is not ontological but phenomenological. Therefore, the body-as-subject is the very same physical object as the body-as-object. Nonetheless, it is experienced in a fundamentally different manner (Thomspson 2007; Zahavi 2005; Legrand 2007a,c,d).

---

[20] See section on The body-as-subject.

By definition, the body-*as-subject* is itself absent-*as-intentional-object*.[21] Interestingly, the different ways in which the body is *absent-as-intentional-object* are associated with different ways in which the body is *present-as-subject*.

More in detail, it is often emphasized that the body-as-object falls into some "*background disappearance*" (Leder 1990, 27) when one directs one's attention to some other objects/projects.[22] Importantly, though, the body absent-as-object has its own form of presence: the body is experienced-as-subject *correlatively* to things in the world perceived-as-object. The paradigmatic example is the experience of the hand-as-touching (vs the hand-as-touched), which is not an object of experience but is experienced as-*subject* correlatively with the object touched (Merleau-Ponty 1945).

The experience of oneself is here specifically bodily, in the sense that the body *itself* is specifically experienced-as-subject. In particular, I experience myself as acting (Legrand 2007d) and while acting I experience my body as imposing on me its constraints, and as allowing the execution of my projects in a way other objects don't. This is the case in at least the following three ways.

First, the subject does what the body can. Not only I impose my intentions on my body but also my body allows my desires ("I want") to take the form of concrete projects ("I can"), in the sense of Merleau-Ponty's notion of "motor intentionality" (1945). Moreover, for these projects to take place in a world of physical objects and constraints ("I do"), I need to conciliate my desires with my body's ability ("I must"). Acting ("I do") involves a dynamic equilibrium between the subject's "I can" and the "I must" imposed by the body-as-physical-object (See also Sartre 1943, 327).

Secondly, the body may accomplish what the subject cannot do intentionally. Consider the following example: "I lie down in bed, on my left side, with my knees drawn up; I close my eyes and breathe slowly, putting my plans out of my mind. But the power of my will or consciousness stops there.… I call up the visitation of sleep by imitating the breathing and posture of the sleeper … there is a moment when sleep 'comes', settling on this imitation of itself which I have been offering to it, and I succeed in becoming what I was trying to be: an unseeing and almost unthinking mass …" (Merleau-Ponty 1945, 189).

Thirdly, the subject does what the body imposes. My body is a physiological organism to which "I must" surrender (How often do I *need* to eat? How long do I *need* to cross the street?). Such bodily constraints are flexible but un-eradicable.

At least in these three ways, being a body-as-physical-object contributes to one's sense of self as subject and agent in allowing and constraining one's intentions and behaviour.

---

[21] This does not rule out the possibility that the body can be experienced both as-object and as-subject.

[22] Such absence-as-object is only *contingent* since, as underlined above, a shift of attention can easily lead these body parts to be taken as-intentional-objects.

In addition to the "background disappearance" of the body, the body part which makes anything present-as-object is *itself* absent-as-object. The paradigmatic example is the eye: "The eye does not appear visually.… Naturally, one would not say that I see my eye in the mirror. For my eye, the seeing qua seeing, I do not perceive" (Husserl 1989).[23] However, it would be a mistake to consider that this absence hides the body in complete darkness. The body-as-subject-in-the-world is *transparent* in the sense that one experiences the world *through* it, but transparency is *not* invisibility (Merleau-Ponty 1945, 345; Legrand 2005). It rather corresponds to a form of bodily-consciousness which goes beyond the body proper, as it corresponds to the experience of the world as disclosed by the body, to the "center of reference which things indicate" (Sartre 1943, 320). For example, when a diver looks at the surface of the water down below, before taking a plunge into the sea, he precisely sees the water surface at a certain distance from the location of his body. In that sense, he experiences the location of the sea as much as he experiences his own location, the difference being that he thematizes the former and not the latter (Legrand 2007b). We see here that even if the body does not figure in the *content* of one's experience, it is not completely absent from experience. Rather, one's body-as-subject-in-the-world is pervasively experienced as it *structures* any experience, by anchoring it to the spatio-temporal location of the experiencer's body (Legrand et al. 2007). In this sense, "external perception and the perception of one's own body vary in conjunction because they are the two facets of one and the same act… every external perception is immediately synonymous with a certain perception of my body" (Merleau-Ponty 1945, 237–8).

The experience of the body-as-subject is thus jointly constituted by two forms of experience: the experience of the subject itself as bodily and the bodily anchoring of the experience of the world. In normal circumstances, rather than being explicitly conscious of our body and of the bodily thickness of our being-in-the-world, we project our gaze directly out towards objects in the external world. Nonetheless, the body-as-subject appears in experience, both as it discloses bodily projects and as the very perspective to which any experience is anchored.

Importantly, experiences of the body as-object and as-subject are not only fundamentally different modes of consciousness; they are also constitutively *intertwined* with each other. Indeed, the body-as-subject is both a physical body which can (at least potentially) be taken as-intentional-object, and a subject who experiences himself as such. This appears clearly in the following experience: "when I touch my right hand with my left, my right hand, as an object, has the strange property of being able to feel too.… I can identify the hand touched as the same one which will in a moment be touching" (Merleau-Ponty 1945, 106; see also Husserl 1989).

Last but not least, the intertwining of experiences of the body as-object and as-subject is itself intertwined with one's sense of self. Indeed, the latter fundamentally involves the experience of being-in-the-world, and this is precisely what is

---

[23] See also Sartre (1943, 304, 316); Compare with Wittgenstein (1921, 5.631–5.641).

provided by the experience of the body-as-subject, together with the experience of the body-as-object as mine/me. What is crucial for understanding bodily-self-consciousness is to acknowledge the distinction between the recognition of the body-as-object as mine and the experience of one's body-as-subject. The former experience is a contingent cognitive accomplishment while the body-as-subject is not explicitly experienced as mine but pervasively and unambiguously as me.[24]

All together, these considerations avoid the opposition between presence-as-object and absence, in favor of the distinction and intertwining of presence-as-object and presence-as-subject. Thereby, they offer a framework adequate to further investigate the (im)possibility of disembodiment. We've just seen that, de facto, the body occupies an important part of one's experiential landscape, not only punctually when one pays attention to one's body, but also pervasively, when one experiences non-bodily events. As such, it importantly participates to the constitution of self-consciousness. Moreover, it follows from these descriptions of bodily experiences that, not only being an experiencing subject, but more in particular being an experiencing *body* is as necessary (hence, as contingent) as experiencing a world: As I cannot act without being acted on (Sartre 1943, 324), I cannot see without being visible (*Id.*, 317). My body "is therefore in no way a contingent addition to my soul; on the contrary it is a permanent structure of my being and a permanent condition of possibility for my consciousness as consciousness *of* the world" (*Id.*, 328).

However, these claims might be questioned on the basis of cases where one's experience of oneself and/or of the world may conflict with one's bodily-consciousness, thereby re-opening the question of whether self-consciousness involves bodily-consciousness necessarily or merely contingently. In what follows, I will address this question by examining clinical cases of purported dissociation between self-, body- and world-consciousness.

Being a Bodily Subject Out of One's Body

As described above (section on Four irreducible bodily dimensions), Out-of-Body Experiences (OBEs) involve the feeling of having left one's physical body and, most often, viewing it from an external perspective. The point above was that OBEs do not involve any *physical* disembodiment. Now, the descriptions of bodily-consciousness in the previous section allow us to emphasize that OBEs do not involve *experiential* disembodiment either, despite the name they are tagged with.

First, during OBEs, the body may be experienced as-object. It is, e.g. looked at, and located among other objects. What is interesting is that this body-as-object is unambiguously experienced as one's *own* body, even though the subject experiences

_____

[24] There is no identification of the body-as-subject, therefore no possibility of misidentification (Legrand, 2006, 2007d).

it from an exceptional perspective, as he might never have seen his body from above before, and might not be able to recognize particular physical features of this body-as-object (Blanke et al. 2004).

Secondly, the subject is, by definition, not characterized by a body-image, be it veridical or illusory. This would indeed turn the subject into an intentional object. Nonetheless, even in OBEs, the subject experiences himself as bodily[25] in the sense that OBEs are associated with vestibular sensations (Blanke et al. 2004): subjects report "feeling as if being above their real body", "rapidly rising higher", "in a horizontal position", "floating at the ceiling".[26] Moreover, what remains throughout OBEs is the experience of a world where one is anchored as a bodily-subject ("I look down"). As argued by Merleau-Ponty "what meaning could the word [down] have for a subject not placed by his body face to face with the world?" (1945, 116).

Given these considerations, OBEs are better described as the "experience of seeing one's body in a position that does not coincide with the felt position of one's body" (Blanke et al. 2004, 256), as this description allows the consideration of the experiences of one's body-as-object and as-subject which both remain vivid in most[27] OBEs. Therefore, these experiences are interesting not because they ground the notion of a disembodied self (Metzinger 2005) but because they involve a dissociation between the body-as-subject and the body-as-object, in a way that is normally not the case: "What make [OBEs] unique is that … you see your own body, and you recognize it as your own, but presently it is not the body as subject…" (Metzinger 2003, 497). The latter, however dissociated from the former, remains experienced as such. Metzinger chooses to describe this state as, at best, only "weakly embodied" (Metzinger 2003, 499) but what OBEs more fundamentally reveal is the bodily nature of the subject, whose strength resists even cases of dissociation from the body-as-object. The crucial point revealed by OBEs is that the subject locates himself where the perspective is anchored (Metzinger 2003, 502), suggesting that one's sense of self is anchored in priority to the body-as-subject rather than to the body-as-object.[28]

In OBEs, the subject does not experience as-subject the body-as-object even when he identifies the latter as his own. OBEs thus show that self-attribution of one's body-as-object can occur without this body-as-object being experienced as the locus of one's self-as-subject. However, this should not be interpreted

---

[25] "I did not feel bodiless", Waelti 1983, 18; in Metzinger 2003, 491. In this particular case, the subject reports to experience himself as an agent deliberately moving through space but this aspect would be too exceptional to be considered as a common feature of OBEs.

[26] Importantly, these data suggest a role of the vestibular cortex in the induction of OBEs (Blanke, 2004; Arzy et al. 2006; Lopez and Blanke, 2007).

[27] The form(s) of bodily-consciousness that would prevail even in cases of "asomatic OBEs" remain to be assessed.

[28] Metzinger (2003, 500) concludes that "the self-as-object is not a strictly necessary condition" but this seems to neglect the fact that a major component of OBEs is the experience of the body-as-object, and the experience of it as mine.

as suggesting that consciousness of oneself-as-subject is not necessary for self-consciousness. It rather interestingly means that the body-as-object and body-as-subject are not necessarily experienced as located in the same place. But it remains hard to conceive of the possibility to recognize a body as mine if I do not hold a perspective on it, i.e. if I'm not a bodily subject experiencing myself as such.

All together, this interpretation of OBEs thus reinforces the claim that these two experiences are fundamentally different, and irreducible to each other.

To conclude this requalification of OBEs as involving a double experience of one's body as-object and as-subject, it is relevant to underline that the malleability of one's bodily-consciousness also arouses interest in cognitive neurosciences. For example, empirical investigations suggest that a manipulation as simple as a synchronous stimulation of your hand and of a rubber hand leads you to mislocate your arm closer to the rubber arm, an experience referred to as Rubber Hand Illusion (RHI, Botvinick and Cohen 1998). This experience has recently been replicated with the whole body (Lenggenhager et al. 2007) and may be considered as opening a new way to investigate OBEs empirically. However, crucial differences between these anomalous forms of bodily-self-consciousness are worth mentioning, as it might provide a way to investigate empirically the determinant of different dimensions of bodily-self-consciousness in a fine-grained manner. Of particular interest, in "Lenggenhagerian" OBEs (Lenggenhager et al. 2007) one experiences one's body-as-object as being in a location where the biological body is not (the same goes for the RHI). Contrastively, in spontaneous OBEs, one experiences one's body-as-*subject* as not located where the biological body is.[29]

Interestingly, another experimental manipulation would also elicit OBEs (Ehrsson 2007), but in still a different manner. In this latter case, synchronous visuo-tactile stimulation of the subject's body and an "illusory body" leads the subject to experience himself as located behind his physical body. In this sense, it seems that this experience reproduces one aspects of the phenomenology of OBEs, i.e. the mislocation of oneself-as-subject. However, the subjects report looking at their body "as if it was another's", thereby contrasting with OBEs where self-recognition of the body-as-object is preserved.

To sum up, in "Lenggenhagerian" OBEs the physical body and the perspective of the subject are experienced as located at the same place (the perspective is held from the location of the physical body), and are dissociated from the location of the body-as-intentional-object. In spontaneous OBEs and "Ehrssonian" OBEs (Ehrsson 2007), the physical body and the body-as-intentional object are experienced as located at the same place, and as dissociated from the perspective of the body-as-subject. In spontaneous OBEs and "Lenggenhagerian" OBEs the body-as-intentional-object is self-attributed. This is not clearly the case in "Ehrssonian" OBEs.

---

[29] This coheres with the authors conclusion that "Because the present illusion was neither associated with overt disembodiment nor with a change in visuospatial perspective, we argue that we have induced only some aspects of out-of-body experiences or rather the closely related experience of heautoscopy that has also been observed in neurological patients" (Lenggenhager et al., 2007, 1098).

194 D. Legrand

(De)constructing One's Bodily-Self

So far, purported cases of disembodiment rather revealed the strength of one's sense of self as bodily. On this ground, the current section will be devoted to the case of anorexia,[30] to further clarify the complex intertwining of self and body, and consider whether the distinction introduced above (between the body-as-object and the body-as-subject) can shed some light on this pathology.

The widespread conception of the body and self as dissociable from each other grounds the view that the anorexic seeks the destruction of her[31] body for the sake of herself: on the ground of a distorted body image (Bruch 1973),[32] she would self-destruct her body as a means to meet her highly self-critical standards,[33] seeking the perfection of a *disembodied mind*. By contrast, the notion of "bodily-self" allows the reconsideration of the anorexic patient as seeking self-preservation. To better understand how she might coherently seek the *preservation* of her bodily-self by operating a bodily self-*destruction*, one needs to rely on the aforementioned distinctions. The claim would then become the following: anorexics seek the preservation of their body-as-subject by destructing their body-as-object.[34]

To clarify this view, let us first consider the role of others. Anorexics are often victimized as being enslaved to the ideal of the slim body that occidental societies heavily convey. However, the social impact on women's body does not concern only their body-*image*. Being anorexic is not seeking social perfection in a too radical fashion[35] and the anorexic does not merely compare (in an obsessive and self-depreciating manner) two body-images: her own and the socially applauded slim bodies. What matters is not only the slimness[36] that is conveyed through these images, but also the *objectivation* of the body that society operates, notably by the very fact of considering the body through its image: what is involved here is "the basic fact of the woman's social existence as the object of the gaze of

---

[30]"Individuals with anorexia nervosa are unable to maintain a normal healthy body weight, often dropping well below 85% of their ideal weight" (Bulik et al. 2005, 52).

[31]In the following I will only refer to women as cases of anorexia are found mostly among women. But note that men would make up approximately 10% of anorexia nervosa (Weltzin et al. 2005).

[32]Her body image is distorted both exteroceptively ("I'm fat") and interoceptively ("I'm full").

[33]"Self-devaluation is the essence of the illness" (Bruch, 1978, 154).

[34]It should be clear that this statement does not imply that it's all there is to anorexia which is obviously a more complex pathology.

[35]"Anorexia nervosa is an extremely complex illness, much more than dieting gone wild" (Bruch, 1978, 94).

[36]"The anorexic's rituals – her food compulsions, her exercise routines, her militaristic regimenting of every aspect of her life, every movement and action of her body – have a goal, a purpose, which only peripherally has to do with being thin to be thin, and has everything to do with changing and modifying the self, while at the same time communicating the "attitude" of the self" (Lester, 1997, 485).

another… [as] a thing that exists as *looked at and acted upon*" (Young 1980, 39). Anorexia would thus be grounded on the fact that any subject may have "the impression that the alien gaze which runs over his body is stealing it from him… in so far as I have a body, I may be reduced to the status of an object beneath the gaze of another person, and no longer count as another person for him.… Saying that I have a body is thus a way of saying that I can be seen as an object and that I try to be seen as a subject" (Merleau-Ponty 1945, 193). As Sartre explains about a shy man: "he is vividly and constantly conscious of his body not as it is for him but as it is for the Other.…This is why the effort … will be to suppress his body-for-the-Other. When he longs "not to have a body anymore", to be "invisible," etc., it is not his body-for-himself which he wants to annihilate, but this inapprehensible dimension of the body-alienated" (1943, 353).

This over-objectivation of the body does not rely only on the idealization of its image. Rather, the feminine body is also taken as an object of (sexual) desire. In this sense, the body is corrupted by others' looks and practices, by others who access the body beyond the subject's control (literally or symbolically). By being denied her control over her own body, the anorexic would lose her body-as-subject in the hands of others; she would be left with only her body as object-for-others. She would live this loss of body-control as a loss of self-control, and refuses violently that the body-as-object-for-others takes over her body-as-subject.

The body is not only objectified in the attitude of others "but the woman herself often actively takes up her body as a mere thing" (Young 1980, 44). The anorexic would radicalize this self-objectification and would thus experience a paralyzing failure at being an autonomous subject, an "all-pervasive sense of ineffectiveness" (Bruch 1973, 222, 254). As a consequence, she heavily relies on others for living her "faked existence" (Bruch 1978, 158), letting others drive her sense of self.[37] At the same time, she needs to (re)conquer her "I can", paralyzed by others' "you must", which is most of the time a self-imposed "I must" (e.g. "I must see the pride in my parents' eyes"). As developed above (section on The body-as-subject), one experiences oneself as subject and agent through the bodily intertwining of "I can" and "I must". In their "blind search for a sense of identity and selfhood" (Bruch 1978, x), anorexics would unbalance this equilibrium, as the physical constraints of her body-as-object take over the efficacy of her body-as-subject.

If this view is correct, anorexics are engaged in an ongoing process of self-construction. How, then, can one understand the anorexic's body-destruction? Starving her body is in fact a "magical" solution for the anorexic patient, as she achieves at least five intertwined aims in one step: (1) by her obsessive, ritualized restriction

---

[37]"You think you are worthwhile only if you do something very special, something so great and dazzling that your parents and other people you care about will be impressed and admire you for being super-special" (Bruch, 1978, 137–138). "I was trying to be somebody my parents wanted me to be, or at least the person I thought they wanted me to be… it was really a self-imposed pressure because he never outwardly asked me… I did the best I could, but I guess it wasn't good enough; I failed in all ways but at least from now on I can try to be the best "me" possible and hope that he'll love me just the same even if I can't meet all his desires for me" (Bruch, 1978, 86).

over her body-as-object, she reinforces her (otherwise contested) position as a body-as-subject who is in control of herself ("I can decide to not surrender to starvation"); by the same token, she (2) diminishes others' control over herself/her body; and (3) diminishes self-objectivation by defeating the "I must eat" imposed on herself by her body-as-object; (4) she tailors a body which fades away from corruption, by the stagnation or regression to a desexualized stage; (5) this body is also built as a self-contained system, as the patient drastically controls all that penetrates it, thereby diminishing its vulnerability to invasion and corruption from the outside.[38,39]

Stuck in this "artificial superstructure" (Bruch 1978, 159) however, the anorexic patient in fact compromises with social conformity: by getting slim, her body also gets closer to the object of (sexual) desires that occidental society idealizes. By sculpting her thin body, she does not make it invisible for the social eyes. Rather, her body screams out both slimness and self-control that society cherishes, and she is craving for such feedback. Understanding anorexia requires understanding this vicious circle, by which the anorexic attitude deeply anchors the patient in the situation she refuses.[40] Presupposing her ability to do otherwise would presuppose her ability to be the free subject who she craves to become, while her often over-conformist childhood both opens the necessity and closes the possibility of her self-assertion. Because of this vicious loop, only the radicalization of her self-control can allow her to achieve her self-realization. But again, this is a no-win situation since this very radicalization turns her self-realization into self-destruction.[41]

All together, these considerations propose to characterize anorexia as an attempt of self-construction thanks to body-control. As such, this characterization primarily relies on a view of the self as bodily: the very possibility of the anorexic attitude is grounded on the bodily nature of the self. Paradoxically (due to the aforementioned vicious circle), the "efficacy" of this attitude requires its radicalization, which itself leads the patient to body-destruction. The crucial point for the present discussion is that anorexia would thus be characterized by a tension between the body-as-subject and the body-as-object, rather than by a tension between one's sense of self and of one's body.

This view has theoretical consequences. It underscores the fact that anorexia is not adequately characterized as an attempt to disembody oneself. The body is in the

---

[38] "In hunger I am King" (Kazantzakis, 1963; in Bruch, 1973, 250). "This line … expresses the essence of the inner problem in genuine anorexia nervosa… the anorexics struggle against feeling enslaved, exploited, and not being permitted to lead a life of their own. They would rather starve than continue a life of accommodation" (Bruch, 1963, 250).

[39] The "efficacy" (for the patient's project) of this process may explain why anorexics often deny their illness for a long period. Moreover, "starvation produces serious physiological disturbances that make the condition self-perpetuating" (Bruch, 1978, 95). Eating disorders would constitute the major contribution to mortality from psychiatric disorders (Bulik et al. 2005).

[40] "thinness, the very path of her liberation, is that which further enslaves her" (Lester, 1997, 487).

[41] Note, however, that her conscious realization of the intrinsic hopelessness of her project may allow the patient to open her way out of anorexia.

foreground of the anorexic preoccupation for herself (to the point of subjugation), preoccupation for her ability to find the right equilibrium between her body-as-subject and her body-as-object. Far from being disembodied, the anorexic patient would thus be radically bodily. She suffers from a distorted body-image but the actual content of one's experience of one's body-as-object is contingent relative to what it fundamentally reveals: the bodily-self discloses an experience of itself both as-subject and as-object (for oneself and others). This subject-object dialectic grounds the possibility of the unbalanced self-construction faced by anorexics.

This view also has clinical consequences[42] as it implies that therapy should not only focus on the restoration of a "correct body image" but in priority "help a patient in her search for autonomy and self-directed identity" (Bruch 1978, 143).[43] However, the vicious loop is again that others' attention reinforces the value that the anorexic gives to the fact of being respect*ed* which, together with the weakness of her self-respecting perspective, reenacts her lack of autonomy.[44]

Conclusion

While anorexia mainly underlines the potential need to fight in order to preserve one's experience of oneself-as-bodily-subject, in a situation where it risks to be outshined by the objectivation of the body, Out-of-Body Experiences illustrate the

---

[42] Even though no biological factors have been mentioned, this discussion does not mean to neglect the importance of therapies developed from the consideration of the biological factors that would contribute to the etiology of anorexia.

[43] The view that anorexia has to do with self more than with body-image per se coheres with cognitive theories (Williamson et al., 2002) and empirical investigations in psychology (Benninghoven et al., 2007) and neurosciences (Seeger et al., 2002) according to which body-image in anorexia would be more distorted in self-directed judgments than in judgment of the body image of an average woman.

[44] Anorexics would thus experience the therapist's interpretations as "indicating that someone else knows what they truly mean and feel, that they themselves do not understand their own thoughts. The goal of individual therapy should be to help them develop a valid self-concept and the capacity for self-directed action… it does not matter whether or not the interpretation is correct; what is harmful is that it confirms a patient's fear of being defective and incompetent" (Bruch, 1978, 130). Likewise, even though it initially aimed to denounce anorexia, recent pictures of anorexic women (e.g. "No Anorexia" by Oliviero Toscani) may well be counter-productive, as they objectify the body which attends to live as subject. By being so projected, her body is "de-selfed" again. Moreover, "we might even speculate that if anorexia nervosa becomes common enough, it will lose one of its characteristic features, the representing of a very special achievement. If that happens, we might expect its incidence to decrease again" (Bruch, 1978, xii–xiii). However, one might add today, this will not happen if anorexic bodies are pictured and projected. Indeed, this advertisement cannot help patients "giving up this unnatural pride in something that doesn't accomplish anything" (Bruch, 1978, 139).

strength of one's experience of oneself-as-bodily-subject, despite its potential dissociation from the body-as-object recognized as "me". All together, these clinical cases underline the specificity of these irreducible forms of bodily-self-consciousness as well as the importance for self-consciousness to involve the experience of one's body as-subject and as-object as intertwined with each other in a dynamic equilibrium.

Neither these cases, nor the theoretical views discussed previously (sections on A certain unity and Four irreducible bodily dimensions), provide any argument in favor of a disembodied view of self-consciousness, but rather underline the necessity to develop a multidimensional consideration of the body (-consciousness) and its interplay with self (-consciousness).

Surely, the failure of the most radical attempts to demonstrate the contingency of the bodily nature of self-consciousness is not as strong an argument as a direct demonstration of its necessity. However, we know that (1) factually, bodily-consciousness is a genuine form of self-consciousness; (2) factually, bodily-consciousness plays a constitutive role in the structure of consciousness; and according to the present investigation; (3) even radical attempts fail to fully eradicate the body. These points at least justify putting the burden of proof on the *shoulders* of defenders of a disembodied view of the self.

References

Arzy S, Thut G, Mohr C, Michel C, Blanke O (2006) Neural basis of embodiment. J Neurosci 26(31):8074–8081

Benninghoven D, Raykowski L, Solzbacher S, Kunzendorf S, Jantschek G (2007) Body images of patients with anorexia nervosa, bulimia nervosa and female control subjects: a comparison with male ideals of female attractiveness. Body Image 4:51–59

Berlucchi G, Aglioti S (1997) The body in the brain: neural basis of corporeal awareness. Trend Neurosci 20:560–564

Blackmore SJ (1982) Beyond the body. An investigation of out-of-body experiences. Heinemann, London

Blanke O, Arzy S, Landis T (2008) Illusory perception of body and self. In: Goldenberg G (ed) Handbook of neurology 88:429–458

Blanke O et al (2004) Out-of-body experience and autoscopy of neurological origin. Brain 127:243–258

Blanke O, Ortigue S, Landis T, Seeck M (2002) Stimulating illusory own-body perceptions. Nature 419:269–270

Blanke O, Thut G (2006) Inducing out of body experiences. In: Della Sala (ed) Tall tales. Oxford University Press, Oxford

Botvinick M, Cohen J (1998) Rubber hands 'feel' touch that eyes see. Nature 391:756

Bruch H (1973) Eating disorders. Obesity, anorexia nervosa, and the person within. Basic Books, New York.

Bruch H (1978) The golden cage. The enigma of anorexia nervosa. Vintage Books, New York

Brugger P (2006) From phantom limb to phantom body. Varieties of extracorporeal awareness. In: Knoblich G, Thornton IM, Grosjean M, Shiffrar M (eds) Human body perception from the inside out. Oxford University Press, Oxford, pp 171–209

Bulik C, Reba L, Siega-Riz A-M, Reichborn-Kjennerud T (2005) Anorexia nervosa: definition, epidemiology, and the cycle of risk. Int J Eat Disord 37:52–59

Chiel HJ, Beer R (1997) The brain has a body: adaptative behavior emerges from interactions of nervous system, body and environment. Trend Neurosci 20:553–557

Cosmelli D, Thompson E (2009) Embodiment or envatment? Reflections on the bodily basis of consciousness. In Enaction (Stewart J et al., eds). MIT Press

De Ridder D, Van Laere K, Dupont P, Menovsky T, Van de Heyning P (2007) Visualizing out-of-body experience in the brain. N Engl J Med 357:1829–1833

Descartes R(1641) Meditations Concerning First Philosophy (trans: Cottingham J 1996). Cambridge University Press, Cambridge

Ehrsson H (2007) The experimental induction of out-of-body experiences. Science 317(5841):1048

Gallagher S (2005a) How the body shapes the mind. Oxford University Press, Oxford

Gallagher S (2005b) Metzinger's matrix: Living the virtual life with a real body. An interdisciplinary journal of research on consciousnesss, Psyche

Gallagher S, Zahavi D (2008) The phenomenological mind: an introduction to philosophy of mind and cognitive science. Routledge, London

Heinämaa S (2003) The living body and its position in metaphysics: Merleau-Ponty's dialogue with Descartes. In Zahavi, Heinämaa, Ruin (eds) Metaphysics, facticity, interpretation: phenomenology in the nordic countries. Kluwer, Dordrecht

Husserl E (1989) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy; second book: Studies in the phenomenology of constitution (trans: Rojcewicz R, Schuwer A). Kluwer, Dordrecht

Iriki A, Tanaka M, Iwamura Y (1996) Coding of modified body schema during tool use by macaque postcentral neurones. Neuroreport 7:2325–2330

Kazantzakis N (1963) Spain (trans: Mims A). Simon & Schuster, New York

Knoblich G, Thornton IM, Grosjean M, Shiffrar M (eds) (2006) Human body perception from the inside out. Oxford University Press, Oxford

Leder D (1990) The absent body. University of Chicago Press, Chicago, IL

Legrand D (2007a) Pre-reflective self-as-subject from experiential and empirical perspectives. Conscious Cogn 16(3):583–599

Legrand D (2007b) Subjectivity and the body: introducing basic forms of self-consciousness. Conscious Cogn 16(3):577–582

Legrand D (ed) (2007c) Subjectivity and the body. Special issue. Conscious Cogn 16(3):577–768

Legrand D, Brozzoli C, Rossetti Y, Farnè A (2007) Close to me: multisensory space representations for action and pre-reflexive consciousness of oneself-in-the-world. Conscious Cogn 16(3):687–699

Legrand D (2005). Transparently oneself. A commentary on Metzinger, being no one. Psyche 11(5)

Legrand D (2006) The bodily self. The sensori-motor roots of pre-reflexive self-consciousness. Phenomenol Cogn Sci 5:89–118

Legrand D (2007d) Naturalizing the acting self: subjective vs. anonymous agency. Philos Psychol 20(4):457–478

Legrand D (2007e). Pre-reflective self-consciousness: on being bodily in the world. Janus Head, Special Issue: The Situated Body 9(1):493–519

Lenggenhager B, Tadi T, Metzinger T, Blanke O (2007) Video ergo sum: manipulating bodily self-consciousness. Science 317(5841):1096–1099

Lester R (1997) The (dis)embodied self in anorexia nervosa. Soc Sci Med 44(4):479–489

Lopez C, Blanke O (2007) Neuropsychology and neurophysiology of self-consciousness. Multisensory and vestibular mechanisms. In: Holderegger A, Sitter-Liver B, Hess CW, Rager G (eds). Hirnforschung und Menschenbild. Beiträge zur interdisziplinären Verständigung. Academic Press, Fribourg and Schwabe, Basel, pp. 183–206

Lunn V (1970) Autoscopic phenomena. Acta Psychiatr Scand 46(Suppl 219):118–125

Merleau-Ponty M (1945) Phénoménologie de la perception. (trans: C Smith, 1962). Tel Gallimard, Routledge Classics, Paris

Metzinger T (2003) Being no one. MIT Press, Cambridge MA

Metzinger T (2005) Out-of-body experiences as the origin of the concept of "soul". Mind Mat 3(1):57–84

Putnam H (1982) Reason, Truth, and History. Cambridge University Press, Cambridge

Sartre J-P (1943) L'être et le néant (trans: Methuen et al., 1958)Tel Gallimard, Paris

Seeger G, Braus D, Ruf M, Goldberger U, Schmidt M (2002) Body image distortion reveals amydala activation in patients with anorexia nervosa – a functional magnetic resonance imaging study. Neurosci Lett 326:25–28

Thomspson E (2007) Mind in life: biology, phenomenology and the sciences of the mind. Harvard University Press, Cambridge, MA

Weltzin T, Weisensel N, Franczyk D, Burnett K, Klitz C, Bean P (2005) Eating disorders in men: update. J Men's Health Gend 2(2):186–193

Williamson DA, Stewart TM, White MA, York-Crowe E (2002) An information-processing perspective on body-image. In: Cash TF, Pruzinsky T (eds) Body image: a handbook of theory, research and clinical practice. Guilford, New York, pp 47–54

Wittgenstein L (1921). Tel (ed 1961) Tractatus logico-philosophicus. Gallimard, Paris

Wittgenstein L (1958) Preliminary studies for the "philosophical investigations". Blue and Brown Books, Gallimard, Paris (1996)

Young I (1980) Throwing like a girl: a phenomenology of feminine bodily comportment, mobility and spatiality. Hum Stud 3:137–156

Zahavi D (2005) Subjectivity and Selfhood: Investigating the first-person perspective. Bradford Books/MIT Press, Cambridge, MA

A Husserlian, Neurophenomenologic Approach to Embodiment

Jean-Luc Petit

A Description of Lived Experience

Normally, we perfectly know where we are: we know that we are viewing the world from a particular and unique point of view. We know that we inhabit the physical body that is situated precisely at this same point of space. We localize ourselves in an absolute manner: definitely, we are 'here'! We know where our hand is without having to watch it constantly, and we move our body without having to look around to know where it went. We can reach out for an object close at hand without having to fix it attentively in advance in order not to miss it. Such 'knowledge' (a misnomer) is indispensable in order for us to deal in a rapid, silent, adaptive and efficacious way with our customary occupations and duties. It is only when anomalies occur linked to cerebral lesions which make unreliable this implicit 'knowledge' and distort our experience that we become aware of the fact that this experience is contingent upon unknown conditions. Conditions that neuroscience help us understand by linking them to dysfunctions of the mechanisms underlying our sense of the moving body: 'kinaesthesia'.[1]

Let's first get straight about the etymology of 'kinaesthesia': *kinêsis* (movement) derived from *kineô* (to move) + *aisthêsis* (sensation, perception) derived from *aiô* (to hear, perceive). Apparently, neither more nor less than a sensation of movement. In the Husserlian theory of kinaesthetic constitution that interests us here, the term 'kinaesthesia' is only employed, for want of anything better, to designate both the lived experience of posture, the orientation of the motor organs of perception

J.-L. Petit
Université de Strasbourg and Laboratoire de Physiologie de la Perception et de l'Action, Collège de France, Paris

[1] Lately in neuroscience there is a growing interest in an extreme disturbance of embodiment: 'out-of-body' epileptic experience, 'the intriguing experience of seeing one's body in a position that does not coincide with the felt position of one's body' (Blanke et al.2004; see Blanke et al. 2005; Blanke and Mohr 2005; Arzy et al 2006).

and movement, even the acts used to simulate these movements from within. The kinaesthetic sense does not limit itself to a proprioceptive, muscular or sinews sense of displacement of the limbs or of locomotion of the whole body contingent upon peripheral stretch captors in the muscles. Above all it implies our more mysterious[2] vestibular sense of inner and outer egocentric space, with its coordinate axes rooted in the body: our sense of our massive limbs and body going forward or backward, up and down, right or left or turning round; our sense of a changing velocity in this movement, our sense of effort, of impulse or resistance against an alien force; not of perspective, but of perspective taking and changing, in the body or mentally only, etc.[3]

This is all part of our blind, preverbal, implicit and immanent knowledge we all exercise in daily life. In going over these aspects of daily life we are not really going beyond a phenomenological description of experience in its constant, structured character. It is interesting to try and bring together what belongs to such description of the lived experience of everyday life and what belongs to the objects of research of the neuro-physiologist. The true foundation of the work of the physiologist is the description and explanation of a repertory of behaviour covering the field of our interaction with the world. The familiar availability of this repertory of behaviour might give us the impression that we benefit from knowledge of the external world that guides our acts, making it possible to accomplish them even when there is no sensorial inputs.

The step that needs to be taken is the one that leads from a phenomenology of everyday life, just as it is lived out, to the theory of the subjective and objective constitution of the experienced world. If we manage to draw up a repertory of those structuring features which confer upon our lived experience its form, if we succeed in tracking these structuring features of experience back to the neural mechanisms underlying them, then we will be in a position to radically elucidate the sense of being of everything that inhabits the world of an agent – including the thing that this agent himself is. Thanks to this theory of constitution all these things would emerge endowed with sense from the fact that we would have identified and distin-guished up to their central underpinnings their organizing principles. Evidently, in saying that, we are only making a tentative gesture in a direction that we would be happy to point to as the right one for the neuro-phenomenologist to go.

Only that, what might leave us unsatisfied from a phenomenological point of view is that one keeps falling back on mechanisms whose objectivated functioning can never really be properly 'internal', that is to say lived from within, because it is only observed from without. Representing processes of the living being through

---

[2]Its brain correlates are still being discussed: parietal, temporal, insular cortex or temporal-parietal junction? (see Lobel et al. 1998).

[3]In the following list of publications, certain aspects of the role of the vestibular system will be found: Berthoz 1973, 1978, 1991, 1994, 1996, 113–125; Berthoz, Graf and Vidal 1991; Berthoz and Melvill Jones 1995; Berthoz, Pavard and Young 1974; Buizza et al. 1979; Israël and Berthoz 1989; Ivanenko et al. 1997.

objective mechanisms whose functioning can not fail to miss the specificity of interior experience remains a permanent danger to which even phenomenology is exposed. In a manuscript from 1931, Husserl undertakes a scrupulous description of the dynamics of kinaesthesia, tracing this dynamic right back to instinctual impulses, in the phenomenon of the orientation of vision and the projection of the hands in the direction of an object of interest: specifically for the new-born baby, the mother's breast, as soon as it distinguishes itself from the indifferent background of experience. After a few pages, haunted by a doubt, Husserl asks himself the question: 'can one adopt this description? For all of this can *take place* without my being attracted to it or turning away from it in aversion.[4]' To avoid this outcome, a theory of constitution has to be developed that makes the agent alone, reduced to the sole resources of its internally felt bodily capabilities, responsible for the act of constituting (giving sense to) the action he is about to perform. 'The act' Husserl insists, 'is in the: I am doing,' I am still active throughout the entire time period in which the act is carried out, and it is I who make this act happen.'[5]

One's Own Body

> This is what distinguishes the body from all external things. On the one hand, the body of flesh and blood is also a thing, a physical thing like any other […]. It is a thing among other things, having its changeable location amidst them […]. On the other hand, this thing is precisely 'my own body' (*Leib*), what upholds my 'I'; the 'I' has sensations and these sensations are localized in the body, in part through thought, in part in a more immediately apparent way' (Husserl 1997, §47)

The body I am in is the body in which feeling, perceiving, knowing, prevails. 'Prevails' translates the German verb '*waltet*' which means: to be active in one's body, act in and through it or occupy it effectively. In any case, prevails in the sense of being there. If one wants to create a place for one's own body in phenomenology, it has to be thought of as having the meaning of being precisely that body where I prevail. This is the body in which I experience my states, my sensations, just as much those that are traditionally known as 'external' – despite the fact that they are, and can only be, internal, as kinaesthetic sensations. In a word, this body possesses a sense, and this sense has to be understood in a dynamic manner, by trying to recover the operations to which it owes this sense.

Here we find ourselves in an extremely tense situation whose description takes on the troublesome form of a tautology, that of having to designate, in the perceiving subject who has the experience of being a body, the constituting operations which have made it possible for this body to acquire the sense of being precisely what it is, an own body. The paradox consists in the fact that the body, as the unity of its

---

[4] E. Husserl, *MS* C 16 I (end 1931), 18.

[5] E. Husserl, *MC* B III 9 (October–December 1931), The Problem of the Act, 19.

organs, does not precede but rather follows upon the constitution of the organs of which it is composed. Clearly, I must first have hands, feet, eyes, etc., and use them practically in my relations with things in order to be able to acquire the sense of being that body whose organs are its parts[6]. As a descriptive problematic, the situation is much tighter than in the case of the perception of an external object, of something that is not me, and on which I only take up a point of view. We will have to come to terms with acts which are: (1) meaning giving, (2) the acts of the perceiving subject itself, (3) the very same acts as those brought into play in order to confer a sense upon surrounding objects.

However, if I want to constitute the body, that is, give the body a sense in the same way that I confer a sense upon surrounding objects when I perceive them, I bump up against an obstacle. For in order to confer a sense upon my own body I have to rigorously apply the same operators that I applied in the case of objects, if only for want of others. In advance, I have to suppose that the object remains the same through the different aspects of it that I catch when I go around it, when I pick it up, a variety of aspects that I have to pull together as a whole, etc., all of which I do as a matter of course. So I have to be able to deploy in perspective a series of adumbrations of this own body, reunite the relevant series of adumbrations, thread them all together as one unitary series through which the thing gets posed as being itself single and identical.

Can't this be done for one's own body? No, because none of these attempts at a description, at meaning giving, succeeds in the case of my own body, on account of the fact that, in this case precisely, '*I am inside it*'. What is it then that distinguishes the body in so special a way? First of all, the body is 'always here'. By contrast, external things are there, or rather can only become things for me to the extent that those that are here can be placed over there. And that their different appearances when they are here and when they are there do not stand in the way of my recognizing them to be the 'same things'. The immanental relation in which I stand with respect to my own body makes it impossible for me to take up such a point of view upon it. Indeed, this is what is meant by immanence: the fact of being one with my body entails the impossibility of taking up with respect to me an other point of view, with a view to developing adumbrations of me. But in the absence of this power of adumbrating, of forming pre-objective varieties that can both be differentiated and integrated in the course of experience in the way that is typical for things, things could not exist for the perceiving subject that I am, and so not with respect to my own self, thing that I also am.

This difficulty can be taken up another way: it has been noted that certain patients do not recognize their body as belonging to them, or attribute to someone

---

[6] E. Husserl, *MS* D 10 III (June 1932), 36: 'Genetically, and in and of itself, one's own body is not constituted ontically prior to the constitution of its moving and changing parts, nor to the practical mastery, by the Ego, of the course of their changes. The constitution of the members of one' own body, as physical contents appearing in the visual field precedes that of the unity of the body; in their reciprocal constitution they get unified as 'organs' of one and the same body.'

else the possession of their arm, for example. Somato-agnosies vary widely, and somato-paraphrenia is one of the examples of this inability to recognize a part of the body as one's own. Schizophrenic patients have difficulty in attributing movements of parts of their body to themselves, a disturbance involving 'agentivity' a deficit probably due to the lack of linkage between areas of the brain producing the movement and those perceiving it (see Frith 1992; Frith et al. 1991; Daprati et al. 1997; Jeannerod 2002, 169–185). Identifying one's body as one's own is an actively cerebral, and therefore eminently vulnerable, process.

Perhaps this obstacle is not insurmountable. One way of understanding this contrast between the manner in which we confer a sense upon our body and the manner in which we confer a sense upon other objects is to draw attention to the fact that objects of perception are all of a certain type, always the same. *Grosso modo*, an object of perception can be manipulated, detached from its context as an object upon which I can focus at will while ignoring the surrounding world. This is evidently not the case with one's own body.

One's own body does not emerge out of a constitution of this kind because it can only be present as the environment of each perceived thing, never as a thing which would in turn have to be situated in an environment providing it with a background. A radical impossibility: the fact of my own immanence vis-à-vis my body excludes the very possibility of taking up perspectival views on it, views which I would only be able to assume if I could get away from it. And this in turn would immediately exclude the possibility of giving a meaning to my body as the intentional pole of unity: the only possible way in which meaning-giving can arise.

But can we stop there and simply refuse to accord any one meaning of being to one's own body? Alternatively and all things considered, do we really need a sense of its unity?

Precisely, Francisco Varela (Varela et al. 1993) steeped in an intellectualist reading of Husserl that I reject, criticized Husserl as a thinker who failed to recognize what seemed to him to be 'a fundamental fact': that the ego, the perceiving I, does not exist as a unity. For his part, Varela gave great weight to a contradiction between the shattering and fragmentation of the perceiving I, going as far as the absence (according to his Buddhist belief) of a unique ego, on the one hand, and, on the other, the intimate feeling we have regarding the unity of our existence as agent, the intimate feeling of the permanence of our identity as body-object, etc.

Delving into Husserl's unpublished manuscripts has given us the right to refute this allegation in what concerns Husserl by affirming the following: whatever may be the case regarding 'the identity of the I and the World', the kind of Buddhist belief that no one could possibly expect to find in this representative of the transcendental philosophical tradition, the tension (if not the contradiction) between a shattered I and a unified I was so little overlooked by Husserl that it doubtless became an existential drama haunting his later thinking.

> What are the kinaesthesiae themselves and what relation do they have with the acts of the I that accompanies them; the I which is directed through them and across them to …? Perhaps this very attempt to talk about acts directing kinaesthesia is already misguided? […] Are kinaesthesiae really something I-like? But then, what does the subjectivity in

question here mean?[7] […] What is the particular affinity between this kinaesthetic process and the I in its activity? […] Is the I itself anything at all outside of its concrete acts in the concretion of actual life[8]? […] But then what are we to make of pure subjectivity, the identical I of affection, of real acts, of feelings? […] The same world for me – the same I. What kind of identity are we talking about? *Perhaps constitution is multifarious*?[9]

Multi-sensorial Integration Through the Act

In this feeling of certainty we have regarding our ability to act we find implied, all the same, a 'certain sense' of the unity of our body. We could not act if we were multiple and if, what is more, persuaded of the fact of being so. Alain Berthoz (2000, 93–96) suggested that autistic children suffer from a deficit in the constitution of the unity of the own body and that it is this shattered perception of themselves that prevents them from establishing a relation with others. Without an own body – Husserl reminds those who get excessively enthusiastic about the empathic fusion of consciousnesses of precisely this – no social communication[10]. Pathological deficiencies in the sense of what it means to be an *I can* may even lead to the commission of criminal acts; to take an example offered by Hegel (1949), when an arsonist says that it wasn't he who set alight the house and eventually the entire town, but his hand, one should not rule out the possibility that he might be sincere (in the case of dissociated identity disorder or anarchic hand) – contrary to what is suggested by the purely institutional and legal (not existential) conception of the origin of responsibility defended by Hegel himself.

---

[7]E. Husserl, *MS* D 10 IV (June 1932), Difficulties with kinaesthesia, 9, 11.

[8]Here is the full quotation: 'What is the specific affinity between this kinaesthetic process and the I in its activity? […] But is the I something over and above its concrete acts enacted in the concreteness of life, and is a concrete act thinkable otherwise than as a process through which something runs off, something that could just as well run off from the self itself, inactively, or again, as a nodal point which runs off in an immediately active way and which is even immediately activable. But this is an originary property, therefore not one which can be immediately activated on one occasion and not on another. In the same way that the I is awake, therefore active as awake, it is always already in its activations. Before all else, it is active in its immediately active processes.' *Ibid.*, 13–14.

[9]*Ibid.,* 18, emphasis added.

[10]E. Husserl 1973, I, 70: 'The aesthesiological-kinaesthetic layer of constitution makes of the body (*Leibkörper*) an own body (*Leib*), a sensorial field and an organ of the Ego. And in such a way that the subject can express itself through this body, have corporeal expressions that serve to express its states of mind. […] If it is true that one can separate the bodily (aesthesiological) and the mental (spiritual), the first is primordial (*das Primäre*) for empathy (*Einfühlung*). It is absurd to connect the whole problem of empathy with purely expressive movements, corporeal expression of the mind, as one ordinarily does, and as Lipps himself also did in his research, however valuable that research might have been. Grasping the 'expressivity' of acts and mental states depends on a prior grasp of the body (*Leib*) as own body (*Leib*).'

That there might be a unity of one's own body despite the multiplicity of referential frameworks, of selections between captors, internal simulations of different parts of our body, remains a central problem that has to be tackled in the contemporary physiology of action. Whether one is talking about *coherence* (Berthoz 2000, Chapter 3) or *binding* (Singer 1990), what has to be understood is effectively the linkage between the incomplete perspectives we have of our body, and which are sequentially ordered in time. How do we come by the idea of the permanence of the body?

This question is not perhaps as overtly thematized in Husserl's unpublished manuscripts, but it is to be found there all the same. And in this sense, that, in order to understand the specific constitution of one's own body across its kinaesthetic adumbrations, a change of category proves necessary. That is, by no longer considering one's own body solely as the result of what is called today 'a multi-sensorial integration' and by the way sticking to sensation, but by crossing over to the category of action, in order the better to understand its sense. Because one's own body as a thing having its own meaning of being is after all something practical, caught up at each instant in an intentional act.

Later on, it might turn out that the body also proves to be an aesthetic thing, an object of theoretical enquiry, or whatever you want. But in the first instance it has to be taken as something practical. What confers a sense upon the body for us, is the fact that it is the original location of our intervention in the surrounding world through our actions – in a word, it is itself made up of our actions. To act is to grasp things and to appropriate these things by making use of them as a function of our intentions. To act therefore presupposes that there is a world already constituted for us, a world of things which already enjoy certain stability, certain permanence insofar as they are arrayed around 'me', which in turn presupposes the prior constitution of these things.

A fruitful way of approaching the problem is through certain later attempts made by Husserl at a phenomenology of the instrument, stemming from the 1930s.[11] The ability to act also includes the possibility of extracting one thing from that system through which it is given, the system of orientations. Something is presented as being close at hand or far off, to the right or to the left, etc. We can tear things away from this system through which they spontaneously present themselves to us and divest them of their external properties, with a view to tying them in to the own body. This moment of linkage with external things which, in the first instance, seemed to be things whose disclosure remained independent of the own body, is the decisive moment for action.

In this way we can selectively associate a thing with the source point, that is, the place where the perceiving subject is situated. We find here two aspects, two presuppositions of action which are also two factors in the constitution of the own body:

1. The fact that there are things whose mode of presentation depends both upon my own accompanying movements and upon my compensating for their own movements in order to keep their unity: visual constitution.

---

[11] E. Husserl, MS C16 V (September 1931); MS D10 I (1932) à III (June 1932).

2. A fact that no longer depends upon the above mentioned visual constitution, that
    is, upon the eyes and the organs supporting the eyes, but rather upon my ability
    to grasp with the hands, a new, *haptic* power.

How does this haptic constitution of my own body proceed? It proceeds by
taking hold, a taking hold through which we appropriate a thing as the extension
of a motor organ. And this taking hold manifests itself, literally, whenever we
manipulate a tool. In an analogical manner, it is operative whenever we put on our
clothing and wear it, whenever we make use of a piece of furniture (the artisan at
his bench), whenever we get into a car and transform it into a vehicle.[12] In every
such instance it is a matter of a linkage that gets set up with the own body and
which depends upon a handling operation. This is what can be extracted from the
notion of action by going back to its corporeal roots.

Husserl supposes that the own body is the result of a double constitution: (1)
The constitution of things that are principally visual; (2) a new constitution which
is typically haptic, that is, tied to the ability we enjoy to take hold of, grasp.
Showing a notable premonition of the scientific knowledge which would have
made it possible for him to recognize the existence of muscular and articulatory
proprioceptors, Husserl insisted upon the tactile aspect of the haptic sense. But it is
evident that he had in view this richer haptic sense, whose power is precisely to
combine the various perceptions, of forces, of cutaneous pressures, of limb
displacements, etc. For example, we pick up a tool. We wear a piece of clothing.
For even wearing is a kind of taking hold of. We make use of tables and chairs in
our daily environment, beginning with our bedroom. We 'run around in a car'. We
'make use of a vehicle'. We travel 'by plane'. All of this results in a description of
the totality of things taken from the point of view of the use we are able to make of
them. What is needed is to re-define the very reality of objects as a function of their
possibility of being linked to one's own body, and in this way, as a function of their
being granted a sense by an agent acting on the basis of his body and exerting
his haptic power, his "hold over …": a sense giving operation which contributes
essentially to this agent making sense of his own body.

Transforming the Subjective into the Objective

For Husserl, we have first of all the system of orientations in the first person per-
spective. One's own body is the zero point on the basis of which directional per-
spectives are unfolded. But in addition to this original role of directed activity it
also captures in a certain way the objects it manipulates and assimilates them to
parts of it. It can therefore also be seen as a sort of compass tracing out the sphere

---

[12]One finds fragments of a phenomenology of changing appearances (points of view, if you prefer)
through the very fact of participating in the movement of a car in Husserl (1997) and of a plane
(or a space ship) in: *Earth does not move,* ms D 17 (in Farber 1940).

of belonging of the ego, and thanks to which I can say of all the things to which I have direct access that they are 'ready to hand'. Among the directly accessible things we find, of course, my body, but also all those objects I make use of regularly, which are more or less directly associated to me.

What then is the relation between these two sets of things? A direct link, as powerful as it is contingent, is created whenever I take hold of an object and displace it, and this by virtue of the fact that it now becomes a part of my body, in a certain sense, and at least as long as I keep hold of it. The constitution of one's own body has a fundamental property: that although it remains invariant it can be modified as a function of the action in progress. An example is given by the temporary extensions of the body implied by the use of tools of all kinds, and in all sorts of circumstance. In his later manuscripts, Husserl insists upon the special status of the act of taking hold of an object in the external world and so making of it a prolongation of the own body. One's own body functions as a transformer of the subjective into the objective and reciprocally. I get into my car. The car becomes an extension of me. I get out, and the car becomes an external object once again. The own body, this curious object, can, at any moment, take advantage of its special status to assimilate to itself an object in the external world, and thus confer upon this relation between the subject and the world the quasi-epistemological status of knowledge through action, knowledge through praxis.

Physiology has to account for a change in the status of the object starting from the moment when it is linked to the body. 'Linked': that could be taken to refer to the desk where I work, to the clothes I wear or, in general, to any extension of the body in the context of a practical activity in which the body is involved. For the latter is caught up in the agent's sphere of belonging. In this frame of reference, what is it then that confers a sense upon the tool? It is the fact that it has become a 'non-kinaesthetic extension of the own-body'[13] as Husserl puts it. Obviously, the tool is not invested with kinaesthetic sensations like an organ of the body, but it participates in the system of the body which is itself constituted kinaesthetically. And this to the extent that the tool is integrated in a transaction at the interface of two systems: the system of the *objectifying perspectives of perception* and that of

---

[13] E. Husserl, MS D10 I (May 1932): 'Non-corporeal things picked out by the place zero do not yield kinaesthetic sensations like limbs, but they participate in such sensations when they are linked to a member (p. 20)'; 'The object grasped in the hand immediately loses its ability to appear at rest or in motion like an external object; it becomes, so to speak, a part of the own body, with the exception of kinaesthesia, which are missing […]. It is precisely this inversion that obviously lies at the root of the possibility of any aperception of the body as own body (p. 58)'. MS D12 I (5 September 1931): 'Even if haptic touching is not yet practical, like pure vision, it has the property of being able to change into pressing, striking, sliding, etc., to the extent that the appropriate pressure is applied. By an appropriate coordination of the fingers employed to touch something from several sides, this touching can also change into handling, carrying, etc. From pure touching in which the image of the *res extensa* is constituted haptically […] a world emerges in which we can intervene by acting, by moving what is at rest, by carrying through changes which were going to happen anyway and, in this way, subjectivizing what, in a certain sense, is simply there, in itself, in external things by including such things within the frame of our own body (p. 34).'

the *ownness sphere of the body*. The tool is in truth only one of the aspects of a more general phenomenological structure of experience referring back to a functional structure of the brain.

In fact this phenomenology of the body, prolonged through the tool, has been quite recently put on the track of its neural correlates thanks to the work of Atsushi Iriki et al. (1996) and his team at Tokyo University of Medicine. They studied the visuo-tactile neurons corresponding to the region of the hand. A monkey carrying implanted electrodes was trained to recuperate food pellets with a rake. The size of the visuo-tactile receptor field for the recorded neurons was measured at three stages of the experiment: before the use of the rake, during, and after. To obtain an evaluation of the visual receptor field of these neurons, a map was drawn up of the activations obtained when, in the immediate environment of the monkey; a food pellet was dragged within his reach.

What was discovered was an extension, in the axis of the rake, of the visual receptor field of the visuo-tactile neurons whenever the monkey used the rake or, in the words of the authors, 'whenever he intended to make use of it'. This receptor field returns to its previous configuration just as soon as the monkey stops using the rake, even if he is still holding it in his hand. Note that this extension does not coincide with eye movements. To prove it, the eye movements were also recorded and the action potentials released by the presentation of food were superimposed upon eye movements recorded while looking. And here we see clearly the disjunction of the two. It now becomes clear that we have here a modification in the inner sense of the own body in action, whose obvious 'cause' is the intentional use made of the rake. The phenomenologist would say that the probable substrate of the kinaesthesia of the hand gets transformed in such a way as to incorporate the instrument employed in the action in progress. The authors were not unaware of what appears to us to be a direct validation of phenomenological description, since they made an explicit reference to Merleau-Ponty.

The Hand Touching and Touched

Everyone is familiar with the peculiar experience that arises when one touches one hand with the other. What happens is this curious ambiguity consisting in one of the hands being perceived as actively touching while the other is perceived as passively touched. A situation that can be inverted at will. We just said that whenever I grab hold of a tool and make of it the prolongation of my body, the mechanism I bring into play contributes essentially to my sense of having a body. But at the same time, let's remember that it also contributes to my body keeping its status as an object. The objectifying dimension of perception – too easily forgotten in the present day fashion of 'embodiment' of perception in cognitive science literature – depends upon this double contribution. A dimension as essential as that which shoves the object away from me, because this object will not be indefinitely associated with my own body. I can always get rid of it. In the same way, the touched hand

remains 'external', becomes an object for me. When it's a matter of the two hands of the same body, the situation turns to be paradoxical: alternatively objectified and subjectively animated, the hands and the entire body are eventually endowed with the sense of a thing, – but a thing that is lived from within!

The attribution of such a sense of being to our own body depends essentially upon kinaesthesia. Husserl distinguishes two categories of kinaesthesia. First, *objectifying kinaesthesia* (KO) functions in the perceptual mode so as to objectify the thing that the touched hand is. This is not, as Husserl reminds us, 'the kind of kinaesthesia that bring our two hands together in accordance with one's desire[14]'. For there is another contribution made by kinaesthesia, a contribution which, this time, goes back up to the source of our motor intentions and which Husserl calls '*motor kinaesthesia*' (KM), and this with a view to distinguishing them from the 'objectifying kinaesthesia'. However we do also *feel* the motor kinaesthesia that invade the touching hand. So, in addition to the feeling of being acted upon, we have to recognize *a feeling underlying action*. In this way, a double kinaesthetic contribution is brought to the theory of constitution. In order to account for the way in which our feeling of being a body arises, we have to bring into play both groups of kinaesthesiae, KO and KM. And if we can play with our different organs in such a way as to evoke this reciprocal touching-touched relation, it is simply because of the contingent, dual way in which the sensori-motor functions of our body have been set up: the touched hand can arbitrarily turn into the touching hand, the touching hand, into the touched hand. Each in turn can be animated by motor kinaesthesia or uphold objectifying kinaesthesia.

We now need to examine the relation between this dichotomy and experiments done on monkeys which show a difference in the activation of the neurons of the superior temporal sulcus, depending on whether its arm is touched by a stick or whether it touches its own arm (Perrett and Mistlin 1990). The neurons of the somato-sensory cortex also exhibit different activities under these two conditions (Nicolelis 1996, 2005; Ghazanfar and Nicolelis 1997). When an object is touched what is made use of is sensorial information to characterize the object, and that has nothing to do with the structures responsible for constituting the own body. On the other hand, when one is passively caressed, then the cerebral activity is principally devoted to the constitution of the own body. With humans this fundamental difference between touching oneself and being touched is brought to light through the tickling experience. We can't tickle ourselves for different regions of the brain are activated when we touch ourselves (which fails to produce the tickling impression) and when we are tickled by others, which does indeed produce the laughing reaction (Blakemore 2003; Blakemore et al. 2000).

How is it then that we come by this feeling that there is a body we inhabit? We come by it because, with our two hands, which are organs of action, we are able to operate haptic links with objects – e.g., a hammer – which thereby become extensions of one's own body and which are arbitrarily substitutable each for the other

---

[14]E. Husserl, ms D 10 III (June 1932), 41.

by virtue of the fact that they have been taken from my environment. Constitution, the attribution of the meaning of being to our own body when we are an agent is linked to the fact that there are motor orders, and also to the fact that there is objectifying kinaesthesia (visual, tactile, proprioceptive afferences). But that is not all. The most important thing is that there is in addition *an interaction and a continual co-evolution between the somato-motor and somato-sensory topographical maps* which make up together a system of re-afferences (a constant returning upon themselves of these afferences) to which we referred under the name 'motor kinaesthesia'. So what has just been said about the hands could be taken up again with respect to other parts of the body.

Of course, the notion of 'maps' has to be taken with a grain of salt, however commonplace its use might have become with the specialists of cortical cartography (Xerri 2003; Xerri et al. 1999). Taken just as it arises through the topographical readings obtained in the anatomical or physiological study of the monkey, this notion certainly does not include what features, in our view, as the neural substrate of the constitution of one's own body: the possibility of *simulating action without moving the body*. This is why the notion of maps has to be understood in an extended sense, following the example of Gerald Edelman (Edelman and Tononi 2000), who conceives them as emerging out of dynamic processes implying entire networks of brain. In the course of ontogenesis and indeed throughout the entire length of individual experience, one process (re-entrance) ensures the selection and the correlating of groupings of cells functionally associated through perceptual activities, even though these groupings are situated in distinct and distant anatomical cerebral regions. This process creates solidarity of mutual linkage between the maps at the root of the perceptual categorization of the environment.[15]

This would mean that there would have to be an integrated system of somato-sensory and somato-motor maps which constantly modify each other, a system brought into play whenever I take hold of an object independent of me, thereby making of it a prolongation of my body, and that it is across this transitory prolongation and shortening again that I acquire the sense of being my own body. Michael Merzenich, to whom we owe the discovery of the plasticity of the sensory and motor cortical maps, writes as follows:

> To a large extent we choose what we will experience, then we choose the details that we will pay attention to, then we choose how we will react based on our expectations, plans and feelings, and then we choose what we will do as a result. This element of choice and the relational nature of awareness in general have almost never been considered in neurophysiological experiments. We realize now that experience coupled with attention leads to physical change in the structure and future functioning of the nervous system. This leaves us with a clear physiological fact, a fact that is really just a mechanistic confirmation of what we already know experimentally: moment by moment we choose and sculpt how our ever-changing minds will work, we choose who we will be the next moment in a very real sense, and these choices are left embossed in physical form on our material selves. (Merzenich and deCharms 1995, 76).

---

[15] Edelman 1989, Fig. 3.1, 45. Other mechanisms have also be taken into account, like those which Rodolfo Llinas (2001) stressed, and which emphasize a continuous oscillatory flux in the loops linking the thalamus to the cortex across the basal ganglia.

Summary

The operations involved in the constitution of the sense of being a body would indeed prove to be a piece of vainglorious conjuring if they were not rooted in our corporeal organization and if they did not bring this organization into play, perhaps in a more eminent manner than ordinary activities. This rootedness in the body is required of any theory of the incarnation of meaning that wants to take up its stand in the wake of lived experience. The act of conferring a meaning upon its body is grounded in the fact that it prolongs, and so makes explicit, the sketch of a corporeal movement founded in the kinaesthetic system.

The neuro-physiological work subtending the constitution of the body could very well not be due entirely to mechanisms of the touching-touched type, that is, the exploration of the own body by itself. Even if the bringing to light of the singularity of one's own body owes a great deal to the philosophically happy choice of this paradigmatic example. In fact, phenomenological description grasps the own body as an acting body, a continually busy body, continually at work at some task. And this is what is meant by the German verb '*hantieren*', which French phenomenologists, following Merleau-Ponty, have rather poetically (and misguidedly) translated as: 'our body haunts the world.' From the physiological point of view, the constitution of one's own body as acting body also, and necessarily, brings into play mechanisms set up to explore the external world.

The fact that one's own body draws its sense of being a body for us from the actions that we are only able to accomplish by bringing into play its practical powers implies that the own body is not enclosed within itself as by a frontier, as is the physical body in the way we ordinarily think of it. It's a lived and not simply a perceived (in the sense of passively represented) body. Accordingly, the diversity of the temporal flux of our intentions, as we project ourselves in all directions foreseeable, is such that there is a constant renewal of the sense of our own body, including the one we have of its geometry. And as long as we are active, the work will not be finished! However, all of this requires – and circularly enforces – an integrated articulation of one's own body, a harmonious synergy as between the different organs, beginning with the two hands.

In the same way, Husserl's robust idea of 'harmony' (his expression for today's physiology 'coherence') was that harmony is always aimed at and that, although it is never wholly achieved, it is constantly sought as a response to a variety of looming dissonances.[16] The vital issue being that of not stumbling over a catastrophic dissonance: Discongruity in the sensorial information available to the perceiving

---

[16] Husserl (1997): 'It belongs to the general essence of conflict, of being otherwise, that it should presuppose a foundation of agreement.' This foundation becomes the horizon of an infinite quest in the later manuscripts where constitution, after being static, becomes dynamic: *MS* B III 9 (October–December 1931) The problem of the act: 'each "creative" real doing serves a universal goal of life centred on the Ego: will to unity, to harmony of being as something that has constantly to be re-established by correction.'

subject – and in the practical intentions of the agent – which might bring with it the destruction of one's own body and its environment, not to say the collapse of the life-world. Something of this kind often does transpire in disturbances like depression, spatial anxiety, agoraphobia (without forgetting out-of-body paroxysmal episodes in epilepsy), but also in such psychiatric afflictions as anorexia, autism, schizophrenia, etc. We are told that the flexibility of the perception of the lived body is such that, in illnesses such as anorexia, certain persons can have the impression that their body is enormous even when they are quite slim and, on the other hand, persons afflicted with elephantiasis can have the impression that the size of their body is normal (see Viaud-Delmon et al. 1999; Viaud-Delmon et al., 2002); Viaud-Delmon et al. 2000). In this regard, physiology and phenomenology both share a commitment in favour of recognizing significance to the anomalies as contributory factors in constituting the world of living that sets them apart from physics. At least to the extent that the (classical) physicist, who works on the principle that he is fully enabled in considering things independently of his access to them, dreams of an exclusively normal system of experience: Never having had any reason to relate sensorial qualities to the corporeal conditions (normal or abnormal) of the subject of experience, this physicist, Husserl observes, 'has nothing to offer the physiologist.[17]

References

Arzy S, Thut G, Mohr C, Michel CM, Blanke O (2004) Neural basis of embodiment: distinct contributions of temporoparietal junction and extrastriate body area. J. Neurosci 2006 26(31):8074–8081

Berthoz A (1973) Contrôle vestibulaire des mouvements oculaires et des réactions d'équilibration. Thèse, Faculté des Sciences

Berthoz A (1978) Vision et système vestibulaire. Encyclopaedia Universalis, Paris, 441–444

Berthoz A (1991) Reference frames for the perception and control of movement. In: Paillard J (ed) Brain and space, Oxford University Press, Oxford

Berthoz A (1994) La Géométrie Euclidienne a-t-elle des fondements dans l'organisation des systèmes sensoriels et moteurs?

Berthoz A (1996) How does the cerebral cortex process and utilize vestibular Signals. In: Baloh RW, Halmagyi GM (eds) Disorders of the vestibular system. Oxford University Press, Oxford

Berthoz A (2000) The brain's sense of movement. Harvard University Press, Cambrigde MA

Berthoz A Melvill Jones G (1985) Adaptive mechanisms in gaze control. Reviews in oculomotor research. Elsevier, Amsterdam

---

[17] E. Husserl, ms D 13 XV (1910–1918), 33. On the importance of anomalies for the constitution of the world of normality. Also see Petit, (1996) *Solipsisme et Intersubjectivité*. Section 2–4, 23–72.

Berthoz A, Petit J-L (2008) The physiology and phenomenology of action. Oxford University Press, Oxford

Berthoz A, Pavard B, Young L (1974) Rôle de la vision périphérique et interactions visuo-vestibulaire dans la perception exocentrique du mouvement linéaire chez l'homme. C.R. de l'Académie des Sciences T.278, D (1974), 1605–1608

Berthoz A, Graf W, Vidal PP (1991) The head–neck sensorimotor system. Oxford University Press, Oxford/New York

Blakemore SJ (2003) Deluding the motor system. Conscious Cogn 12(2003):647–655

Blakemore SJ, Smith J, Steel R, Johnstone CE, Frith CD (2000) The perception of self-produced stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. Psychol Med 30(5):1131–1139

Blanke O, Mohr C (2005) Out-of-body experience, heautoscopy, and autoscopic hallucination of neurological origin: Implications for neurocognitive mechanisms of corporeal awareness and self consciousness. Brain Res Rev 50:184–199

Blanke O, Landis T, Spinelli L, Seeck M (2004) Out-of-body experience and autoscopy of neurological origin. Brain 2004(127):243–258

Blanke O, Mohr C, Michel CM, Pascual-Leone A, Brugger P, Seeck M, Landis T, Thut G (2005) Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. J Neurosci. 25(3):550–557

Buizza A et al (1979) Otolothic acoustic interactions in the control of eye movement. Exp Brain Res 36:509–522

Daprati E et al (1997) Looking for the agent. An investigation into consciousness of action and self consciousness in schizophrenic patients. Cognition 65:71–86

Edelman GM (1989) The remembered present. A biological theory of consciousness. Basic Books, New York, Fig. 3.1, 45

Edelman GM, Tononi G (2000) Comment la matière devient conscience. Odile Jacob, Paris

Farber M (ed) (1940) Philosophical Essays in Memory of E. Husserl Harvard University Press, Cambridge, MA

Frith CD (1992) The cognitive neuropsychology of schizophrenia. Lawrence Erlbaum, Hove

Frith CD et al (1991) Willed action and the prefrontal cortex in man: a study with PET. Proc R Soc Lond 244:241–501

Ghazanfar AA, Nicolelis MA (1997) Non linear processing of tactile information in the thalamo-cortical loop. J Neurophysiol 78(1):506–510

Hegel GWF (1949) Philosophy of right (trans: Knox TM). Clarendon, Oxford

Husserl E (1973) Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. I, Husserliana XIII. Martinus Nijhoff, Den Haag

Husserl E (1997) Thing and space. Lectures from 1907 (trans. & ed Rojcewirz R) Kluwer, Boston, MA

Iriki A et al. (1996) Coding of modified body schema during tool use by macaque postcentral neurones. NeuroReport 7:2325–2330

Israël I, Berthoz A (1989) Contribution of the otoliths to the calculation of linear displacement. J Neurophysiol 62:247–263

Ivanenko YP et al (1997) The contribution of otoliths and semicircular canals to the perception of two-dimensional passive whole-body motion in humans. J Physiol (London) 502:223–233

Jeannerod M (2002) La nature de l'esprit. Odile Jacob, Paris

Llinas R (2001) I of the Vortex. MIT Press, Cambridge, MA

Lobel E, Kleine J-F, LeBihan D,Leroy-Willig A, Berthoz A (1998) Functional MRI of galvanic vestibular stimulation. J Neurophysiol 80:2699–2709

Merzenich M, deCharms RC (1995) Neural representations, experience, and change. In: Llinas R, Churchland P (eds) Mind and brain. MIT Press, Cambridge, MA

Nicolelis MA (1996) Beyond maps: a dynamic view of the somatosensory system. Braz J Med Biol Res 29(4):401–412

Nicolelis MA (2005) Computing with thalamocortical ensembles during different behavioral states. J Physiol 566(Pt 1):37–47

Perrett DI, Mistlin AJ (1990) Visual and somatosensory processing in the macaque temporal cortex: the role of 'expectation'. Exp Brain Res 82:437–450

Petit J-L (1996). Solipsisme et Intersubjectivité. Quinze Leçons sur Husserl et Wittgenstein, Cerf Harvard University Press, Cambridge, MA

Singer W (1990) Search for coherence: a basic principle of cortical selforganization. Concepts Neurosci 1:1–26

Varela F, Thompson E, Rosch E (1993) L'Inscription corporelle de l'esprit. Sciences cognitives et expérience humaine. Seuil, Paris

Viaud-Delmon I et al. (1999) Anxiety and integration of visual vestibular information studied with virtual reality, Biol Psychiatry 47:112–118

Viaud-Delmon I et al. (2000) Adaptation as a sensorial profile in trait anxiety: a study with virtual reality. J Anxiety Disord 14.6:583–601

Viaud-Delmon I, Berthoz A, Jouvent R (2002) Multisensory integration for spatial orientation in trait anxiety subjects: absence of visual dependence. Eur J Psychiatry 17.4:194–199

Xerri C (2003) Plasticité des représentations somesthésiques et illusions perceptives: le paradoxe du membre fantôme. In: Petit J-L (ed) Repenser le corps, l'action et la cognition avec les neurosciences. Intellectica 1–2(36–37):67–87

Xerri C et al (1999) Representational plasticity in cortical area 3b paralleling tactual-motor skill acquisition in adult monkeys. Cerebral Cortex 1999(9):264–276

Body and Movement: Basic Dynamic Principles

Maxine Sheets-Johnstone

Introduction

Embodiment is to the body as enaction is to movement.[1] In each instance, the primary term of the analogy attempts to embrace an animate reality in a way compatible with the science or strand of philosophy being practiced. In each instance, however, the primary term is uncongenial to the basic reality it aims to capture and describe. The lack of fit is indirectly but substantively attested to by indices of books on embodiment and enaction: either no entry exists for the tactile-kinesthetic/affective body and kinesthesia or paltry entries exist. In effect, the foundational ontological and epistemological reality of life is missing: *animation* is nowhere on the map.[2] The lack of fit and missing reality are furthermore attested to by the terms in which proprioception is discussed and the fact that a clear-cut distinction and substantive understanding of the difference between proprioception and kinesthesia is nowhere in evidence (see Sheets-Johnstone 1999 for more on this topic). Proprioception is, properly speaking, not a "matter of debate among philosophers" – seemingly, a matter of determining the correct

M. Sheets-Johnstone
Department of Philosophy, University of Oregon, Eugene, OR 97403-1295

[1]This essay concentrates attention on embodiment. An essay titled "Animation: An Essential, Fundamental, and Properly Descriptive Concept" concentrates attention on enaction and is forthcoming 2009 in *Continental Philosophy Review*.

[2]Opportunities for putting it squarely on the map clearly exist. Consider, for example, Hanna and Thompson's "animalist" solution to what they term "the mind-body-body problem," the latter referring to the tripartite sectioning of what they describe as "something" that is "at once a conscious subject, a living and lived body, and an objective material thing" (Hanna and Thompson 2003b, p. 24). In particular, Hanna and Thompson propose that mind and physical body are simply "*dual aspects* of one's lived body" (ibid.), and that lived bodies are *animal* bodies. Their appeal to "*dual aspect animalism*" (ibid., p. 30) is oddly devoid of any reference to animation, precisely that which would ground their solution in the empirical realities of animate life through reference to proprioception and kinesthesia, the tactile-kinesthetic body and coordination dynamics, and that would flesh out their reliance on "cognitive ethology" (ibid., pp. 31–32) as an academically reputable support for their animalist solution.

answer to a multiple-choice question.[3] Properly speaking, proprioception is a matter of all manner of bodily organs that sense movement and deformations, a primordial form of animate awareness that began its evolutionary career in surface recognition sensitivity – tactility in the service of movement – that evolved into different external sensors registering movement – chordotonal organs, hair plates, sensilla, cilia, and so on – and that, with the advent of internal bodily organs sensing movement through muscular effort, evolved into kinesthesia (Mill 1976; Laverack 1976; Wright 1976; Dorsett 1976; see Sheets-Johnstone 1999 for a close examination and study of the data). As is evident, proprioception is the broader term with respect to kinesthesia. It refers to a sense of movement and position that includes tactility and gravitational orientation through vestibular sensory organs as well as kinesthesia. As its etymology indicates, kinesthesia in its primary, that is, experiential, sense denotes an awareness of *movement*, hence an awareness of *dynamics*, hence an awareness of *a qualitatively felt kinetic flow*. The flow may be felt as smooth, expansive, abrupt, attenuated, jagged, linear, curved, constricted, slow, and so on, including any and all possible combinations as the flow unfolds. Given the inherent qualitative spatio-temporal-energic character of kinesthesia, it is hardly surprising that discussions of body and of movement that omit kinesthesia from their register omit the very stuff of life and the qualitative nature of that stuff. They omit animation.

Understandings of body and movement that are grounded in the natural history of animate life begin with proprioception, with the beginning dynamics of life itself in surface recognition sensitivity, and thereby proceed naturally to understandings that encompass kinesthesia, affectivity, cognition, and the world, including a world of others. They encompass these aspects *naturally* because animation – the dynamics of life itself – naturally engenders kinesthesia, affectivity, cognition, and the world. Movement is in other words at the heart not only of being alive but of staying alive. In an existential as well as evolutionary sense, survival is a matter of effective movement, which means movement that is affectively and cognitively responsive to an ever-changing world that is not the same from 1 day to the next and that demands attentiveness in precisely the way an ant, a spider, a fly, or a human is attentive, not only to the expected and familiar, but to the unexpected or the unfamiliar, the ant, spider, fly, or human recognizing that what is out of the ordinary may perhaps be harmful. As Darwin noted on the basis of his lifelong studies of animate life, "Animals may constantly be seen to pause, deliberate, and resolve" (Darwin 1981 [1871], p. 46).

The moral to be drawn from natural history is that the joints at which humans carve are not necessarily the joints of nature. Artificial joints can indeed give rise to conceptual arthritis in the sense of enlarging the significance of a part, hardening it, and distorting the structure of the whole. They can in turn give rise to linguistic surgeries and therapies that attempt to sew the whole back together in something approximating

---

[3] See Thompson 2007, p. 464, n.3. Thompson cites José Bermúdez's, Dorothée Legrand's, and Shaun Gallagher's "arguments" (as Thompson puts it) as to what proprioception is, that is, whether it is equivalent to prereflective self-consciousness or not and whether the latter consciousness is a perceptual or non-perceptual experience, encapsulating one's body as object or as subject.

its original, wholly natural holistic form. With respect to these linguistic stop-gap measures and operations, researchers would do well to heed the cautionary advice of the Eleatic Stranger from the beginning of their labors. The Stranger cautioned the Young Socrates, "we certainly should divide everything into as few parts as possible" (Plato *Statesman* 287). Certainly if one does not follow the Stranger's advice from the beginning, one should at least insure that the integrity of the whole is preserved, which means that parts are not only put back together, but in a manner that both illuminates and is true to the foundational integrity and nature of the whole.

The moral from natural history and the lesson from Plato are both well illustrated by a fundamental concept in Husserl's writings. Husserl wrote of action, but he did not write of active or enactive organisms; he wrote of bodies, but he did not write of embodied organisms. He wrote of *animate* organisms. Animation is the ground floor, the ontological as well as epistemological bedrock of human self-understandings and indeed of human pan-animate understandings, understandings that include but do not separate out cognition as the point of entry to those understandings.

Embodiment

Present-day cognitive scientists as well as many researchers within both analytic and phenomenological philosophy are wedded to embodiment and its ready-to-hand derivatives – for example, "embodied experience" (Gibbs 2006), "embodied self-awareness" (Zahavi 2002), "embodied subjectivity" (Hanna and Thompson 2003a; Zahavi 2005), and so on. The marriage is one of convenience rather than nuptial depth, and this because cognitive scientists and analytic philosophers intent on profiting from phenomenology and phenomenologists intent on profiting from cognitive science would do well to be studying and describing in detail "the comet's tail of nature" (Husserl 1989, p. 350; on the latter topic, see also Sheets-Johnstone 2007). Fleshing out our foundations in nature requires attention to our natural history, both ontogenetic and phylogenetic. In a – word, it requires attention not to *embodiment*, but to *animation*, for it is in and through animation that we realize ourselves as living beings. We do not come into the world *embodied*. We come into the world *moving*; we are precisely not stillborn. We are indeed animated in basic ways concordant with other forms of animate life, forms whose daily rituals also include eating, sleeping, and mating, and whose affective relations with others and whose cognitive acuities are also central to their well-being.

Cognition is a dimension of *animation*, hence a dimension in the lives of animate organisms. The faculty did not somehow become "embodied" in and with humans – a unique, even *deus ex machina* feature, as it were – but runs the gamut of evolutionary forms of life. In Husserlian terms, not just humans but virtually all members of the Animal Kingdom, a biological category, *turn toward* (or away from) objects and other beings they find in their surrounding world. They are *receptive* (or non-receptive) of them, moving in ways concordant with the meanings those objects and other beings hold for them. More finely put, they move in ways concordant with their affectively

motivated and informed sensory-based cognitions. Members of the Animal Kingdom are indeed *animate* organisms in the full sense of animation, being attuned affectively, cognitively, and kinetically to the world around them. Once cognition is rightly recognized as being both an inherent and integral dimension of the fundamental reality of being alive and moving about effectively, efficiently, and intelligently in the world, there is no doubt that the word *animation* properly describes the bodily nature of cognition.[4]

From the vantage point of animation, three critical lacunae are discernible in both present-day cognitive studies of 'embodied cognition' and its cognates, and in present-day phenomenological studies of 'embodied subjectivity' and its cognates: an attention to kinesthesia and its relationship to fundamental human concepts; an attention to coordination dynamics within the ontogenetical purview of learning one's body and learning to move oneself; an attention to evolutionary biology and its relationship to the coherency or 'existential fit' of *Leib* and *Körper* (see Sheets-Johnstone 1986 on the latter topic). In no instance is it possible simply to 'add and stir', nor is it a matter of 'bridge-building'. On the one hand, first- and third-person perspectives are not simply essentially different perspectives. A first-person perspective necessarily precedes the taking of any third-person perspective: short of first-person experience, there would be no subject or object upon which one could take a third-person perspective. On the other hand, building bridges between first- and third-person perspectives is a hazardous enterprise. It can result in a transmogrification of one of the perspectives, the labor of bridge-building being not necessarily a labor of love and mutual understanding, but one aimed at exploitation if not conquest.[5] Moreover if "Nature is there from the first day" (Merleau-Ponty 1963, 1968; see also Sheets-Johnstone 1999, pp. 306–307), then it behooves us to inquire what precisely is there from the first day rather than either attempt linguistic conversions of Husserlian epistemological phenomenology into, for example, a Merleau-Ponty-based theoretical ontology of "sensible reversibility" (Stawarska 2003; see below) or into a theoretical neurology on the order of "sensorimotor profiles" (Nöe 2004), or attempt linguistic compressions of first- and third-person studies into a hybrid entity on the order of "naturalized phenomenology" (Petitot et al. 1999) or "neurophenomenology" (Varela 1996, 1999; Thompson 2007), or indeed, attempt a mechanics of life featuring sensory inputs and motor outputs far distant from the dynamics of living Nature (Hurley 1998). Though writing in a quite different context, phenomenologist Robert Sokolowski comes close to articulating the starting point for a foundational inquiry into Nature when he observes, "There is no basic consciousness without being awake," a "basic dat[um]" for

---

[4] An editorial concern that the word animation "does not express the bodily nature of cognition" is thus answered; no 'embodiment' needed.

[5] "We have chosen to take as a guideline the idea… that a successful scientific theory of cognition must account for *phenomenality*, that is, … for the fact that for a whole set of cognitive systems, and for the human one in particular, things have appearances. We will argue that on the basis of its past achievements in describing such phenomenality, Husserlian phenomenology can play a key role in helping to meet this requirement, provided that it can be naturalized, and even though Husserl himself strongly opposed naturalism. By 'naturalized' we mean integrated into an explanatory framework where every acceptable property is made continuous with the properties *admitted by* the natural sciences" (Roy, Petitot, Pachoud, and Varela 1999, pp. 1–2; second italics added).

which "[w]e have to thank our bodies" (Sokolowski 1972, p. 76). Indeed, our basic natural rhythm of wakefulness and sleep is a bodily phenomenon. To be at either pole – or anywhere in between – is to be alive – "alive and in the flesh and part of the living, incarnate cosmos," as D.H. Lawrence once wrote (Lawrence 1932, p. 200) – and to be alive is first of all to be *animate*. Even in sleep, we not only move in breathing, but we roll over, bend a knee, extend an arm overhead, stretch a leg, pull the covers up or fling them off. We in fact not only come into the world moving but we go out of the world *un*moving: we are no longer animate in the least part; we are precisely *still*.

Pointed attention to each of the three lacunae will illuminate the phenomenon of animation from a distinctive and crucially integral perspective. Moreover consideration of each in turn will disclose essential aspects of cognition and at the same time show how linguistic and conceptual malpractices hide the very phenomenon that many present-day cognitive scientists and phenomenologists seek to elucidate in their efforts to *embody*, oftentimes gratuitously and even to the point of tautology, not just cognition, cognitive science, and the mind but seemingly every topic of interest (save one)[6]: action (Gibbs 2006), experience (Gibbs 2006), a first-person perspective (Zahavi 2005), subjectivity (Zahavi 2005), practice (Toombs 2001), simulation (Gallese 2007), perception (Gibbs 2006), self-consciousness (Gallagher and Varela 2003), agents (Roy et al. 1999; Varela 1999), intended goals (Gallese 2001), and so on. To term something 'embodied' is akin to anointing it with an ontological salve. The salve putatively binds together mind and body, "the physical" and "the lived," or a first- as opposed to a third-person perspective on humanness. The term itself oftentimes appears gratuitous because the very phenomenon it modifies – for example, agent, action, experience – is already a corporeal-kinetic reality. Indeed, it appears at times tautological: it is as if the body is "embodied." Moreover when we put all features of life that are "embodied" together, we fall far short of an elucidation of human life, and in fact can end up with a meaningless formal declaration on the order of: "embodied agents, through their embodied sensory-motor systems and in their embodied practices, have embodied experiences that they can speak of from an embodied first-person perspective grounded in an embodied subjectivity and an embodied self-consciousness."

Kinesthesia and Fundamental Human Concepts

To begin with, in order to arrive at veritable understandings of kinesthesia and the fundamental concepts generated in and through movement, embodiers need to wean themselves away from sensory-*motor* talk and work toward languaging the

---

[6]To my knowledge, neither cognitive scientists nor phenomenologists have written of "embodied" emotion, an odd omission in their conjoint program of 'embodiment' since emotions are commonly labelled 'mental states'. Social psychologists and anthropologists have, on the contrary, embodied emotion. See, for example, Lyon and Barbalet 1994; Niedenthal et al. 2005, the latter only in terms of showing how "embodiment is critically involved in information processing about emotion" (p. 192).

realities of sensory-*kinetic* experience. This shift involves a shift toward *thinking in movement* (Sheets-Johnstone 1999, Body and Movement: Basic Dynamic Principles), a consistent everyday dimension of animation from infancy onward, not only in reaching and grasping something at hand or in weaving one's way amidst a throng of people on a crowded sidewalk, but in calculating the distance and time to drive from one place to another or in judging the force necessary to splitting a piece of wood. Everyday human experience involves thinking in movement; the everyday experience of animate forms involves thinking in movement.

Our capacity to think in movement is rooted in fundamental human concepts of space, time, and energy or force, all of which are rooted in the experience of movement itself, that is, in kinesthesia. It is of interest to point out in this context an introductory remark in a textbook on movement, specifically a chapter titled "Proprioceptors and Their Associated Reflexes." The authors state, "The voluntary contribution to movement is almost entirely limited to initiation, regulation of speed, force, range, and direction, and termination of the movement." Granted that the authors' approach is neuro-scientific and that their focal interest is in reflexes, still it is striking to read of the "limited" contribution of voluntary movement: not only are the inherently qualitative aspects of movement that constitute its experienced dynamics minimized, but what one might call the very character of living movement in its reality – its initiation and termination – are trivialized. In effect, not only are the felt qualitative dynamics of movement passed over – of which more below – but an integral affective dimension of voluntary movement is passed over: there is no acknowledgement of the motivations that lead to voluntary movement and that inform voluntary movement every step of the way, so to speak, not only in going to the refrigerator because one is unpleasantly hungry or to walk hesitantly to a meeting because one is leery of the agenda, but to rise determinedly out of bed in the morning to go to work in spite of feeling tired.

Speed, force, range, and direction of movement are aspects of movement that may be more finely described phenomenologically (for a full account, see Sheets-Johnstone 1966 [1979, 1980], 1999). Such a description elucidates the complexity of movement beyond the everyday natural attitude toward movement, an attitude that consistently involves wayward notions of movement, to wit: movement is a change of position; we have sensations of movement; movement takes place *in* space and *in* time. A phenomenological investigation of movement discloses qualitative dimensions of movement that testify to movement being a dynamic phenomenon, and being a dynamic phenomenon, it is: (1) falsely defined as a change of position, (2) falsely specified in terms of sensations, and (3) inaccurately described as simply taking place in space and in time. A brief summary of each quality will attest to the falsity and inaccuracy of the above notions and to the complexity of movement. It should be noted that the phenomenologically-disclosed qualities are separable only analytically; that is, they are always integral parts of a whole kinetic dynamic.

Tensional, linear, areal, and projectional qualities are qualities apparent in any movement – reaching for a glass, picking up keys on the way out the door, brushing one's teeth, standing up, sitting down, gesturing in concert with speaking, speaking

itself, and so on, throughout any ordinary day in Western life at least.[7] Being dynamically engendered, each quality is part of a total qualitatively felt dynamic. Tensional quality specifies the felt intensity of a movement, an intensity that may well change in the course of the movement, as in swinging a bat or a golf club to hit a ball, or simply in skipping. The felt and commonly shifting intensities of the movement constitute what is commonly termed its 'force'. Tensional quality captures the felt dynamics of the movement more finely, however, not only in its recognition of shifting intensities, but in terms of its recognition of varying tensional qualities felt through the body in the process of moving, precisely as in the preparatory backswing of a leg prior to its kicking forward and the coincidence of the backswing with an inhalation of breath, for example.

Linear quality describes both the linear design of a moving body and the linear pattern of the movement itself. Both aspects are obviously spatial in character. The linear design of a moving human body might be most readily described in the course of everyday life as upright. That uprightness, however – that *verticality* – not only changes as the result of sitting down, but constantly shifts in the course of everyday walking, for example, when legs are bending and arms are swinging, bringing in diagonal and quasi-horizontal dimensions to what is taken as the vertical line of the body. The ever-changing linear design is indeed part of a total body-in-movement *dynamics*. The same is true for the linear patterns created by movement itself. When we swing our arms back and forth, for example, the path of the movement traces a slightly curved line at our fingertips; the linear pattern traced by each of our feet when we walk traces a more complex line that comes close to describing a flattened circle. When we put the two paths together, precisely as in walking in a relaxed and easy manner, the created contralateral linear patterns of arms and legs attest to the complexity of the dynamics of movement. More complex still are the patterns described by arms moving forward and back, feet circling, and one's whole body moving up and down in jogging. The composite of lines that movement creates can indeed be surprisingly intricate.

Areal quality, like linear quality, has two aspects that again are obviously spatial. They derive from the moving body and from movement itself, areal design describing the former, areal pattern describing the latter. In quite general terms, the areal design of a moving body may be anywhere from constricted to expansive, its shape at the one extreme being small and inwardly oriented, at the other extreme being large and outwardly oriented. Similarly, the areal pattern of a movement may be generally described as anywhere from intensive to extensive, the spatial amplitude of the movement itself being anywhere from small to large. When we are contrite, we tend to shrink in size and stay put, the areal design of our body being small and the areal pattern of any movement we might make being equally small. In contrast, when we run down the street with open arms to greet someone, the areal design of our body is expansive, the areal pattern of our movement extensive. Again, in the context of

---

[7] The qualities of movement do not change according to where one lives or in which culture one is brought up. What does change are the appurtenances that are or are not part of one's everyday life: toothpaste, water from a faucet, a car, and so on.

any movement, areal design and pattern may be a composite, that is, design and pattern alike may be anywhere and go anywhere from one extreme to the other. In the course of sneaking up to grab something while no one is looking and then running away with it, for example, a child – or a thief – may be bent over and take small steps toward the object, then expand sizably along with taking larger steps in running away with it. The areal pattern of a movement may similarly extend to both extremes, as when a person in baseball on first base takes a few steps toward second, retreats, then dashes forward again to second and even around to home.

Projectional quality is apparent in the manner in which movement unfolds, the way in which tensional quality is kinetically manifest. Generally speaking, three different qualities are possible: abrupt, sustained, or ballistic. Infinite degrees of shading are possible within these basic qualities. Moreover, a movement may be a combination of the qualities as when one abruptly moves to catch a glass before it falls and proceeds to place it carefully on the table. While it is sometimes thought that an abrupt movement is always vigorous and explosive and a sustained movement always languid and delicate, such is not always the case: an abrupt movement may be weak, as when eyebrows go suddenly upward and one catches one's breath in moderate surprise; a sustained movement may be strong, as when one pushes a heavy box across the floor.

However brief the above delineations, it should be evident that fundamental concepts of space, time, and force derive from movement, that the concept of direction is rooted in linear quality, distance in areal quality, effort in tensional quality, and so on. The dynamic qualities of movement at the foundation of fundamental human concepts may in fact be exemplified at finer levels. The concepts of 'near' and 'far', for example, are not in the beginning a matter of measured distance at all. They derive from the areal quality of movement; they are rooted in the experience of something being within or out of reach, something being literally 'handy' or something demanding extended movement in order to be attained. Similarly, the concepts of 'weak' and 'strong' are rooted in the tensional quality of movement; they are again not a matter of something measured but a matter of the felt intensity of movement, a whimper, for example, that develops into progressively stronger, ongoing, and resounding wails and cries that engage the entire body in a crescendo of movement and movement-made sound. Clearly, however hidden away in textbooks, neglected in academic discourse, ignored in medical assessments of developing infants and young children, and in general overlooked entirely by people in education, kinesthesia is foundational to fundamental human concepts that develop early on and continue to inform the lives of humans ever onward.[8]

---

[8] It should perhaps be noted in this context that Mark Johnson's anchorage of 'the body in the mind' (Johnson 1987) shows how meaning is generated in and through bodily experience. His analyses are thus in a sense compatible with the basic dynamic principles outlined in this chapter. The compatibility is limited because foundational aspects of animation – kinesthesia, the tactile-kinesthetic body, the developing coordination dynamics of infancy, and the coherency of *Leib* and *Körper* – do not enter into the picture. Johnson's primary concern is language and in fact basic elements in his analyses – "image schemata" and "imagination" – are "embodied." In his most recent book with George Lakoff (Lakoff and Johnson 1999), "embodiment" figures even more strongly.

In sum, it is evident that (1) movement is not a change of position, but the dynamic reality of the kinetic change itself; (2) movement is a matter not of sensations but of a felt qualitative dynamic whose spatial, temporal, and force aspects are the spawning ground of fundamental human concepts; (3) any movement creates its own time, space, and force, and thereby its own particular dynamic.

Coordination Dynamics: Learning One's Body and Learning to Move Oneself

One can see further why kinesthesia – the experience of one's own movement – is not a matter of sensations, but of dynamics. Sensations are spatially pointillist and temporally punctual (for a full account, see Sheets-Johnstone 2003, 2006). Kinesthetic experience is in contrast an experience of an indivisible dynamic whole, a kinetic form that is an overall bodily-kinetic dynamic (for more on this contrast, see Sheets-Johnstone 2003, 2006). Kinesthetic experience is thus not like an itch or a chill or a throb.[9] Precisely with respect to its inherent qualitative integrity and flow, kinesthetic experience is not reducible to a series of before, now, and after moments on the order of one sensation following another as in a series of sharp pains. There is nothing inherent in the series that links the sensations together, certainly nothing on the order of the qualitative dynamics of kinesthetic experience or of the muscular innervations and denervations that neurologically constitute the dynamics, whether a matter of sweeping the floor, getting into a car, or reaching for and picking a book off a shelf. A flow of movement may be accentuated in various ways by shifting intensities or by shifts in direction, for example, or be qualitatively inflected in other ways as indicated above, but the flow is nonetheless coherent, precisely as when one picks up an apple, brings it to one's mouth, opens one's mouth, and bites into it. In a word, the differentially accented and directed flow – *kinesthetically felt movement* – is all of a piece and is experientially and neurologically so constituted.

As so constituted, the flow – the coordinated dynamic – is the basis of both habit and style, the latter a kinetic quality that is the social corollary of one's own personal habits. We recognize style in others and not in ourselves precisely because we are not commonly attuned to our own dynamics. Our attention is commonly consumed elsewhere. We may recognize another person's style outright or implicitly not only in his or her walk, but in his or her laugh or in the way he or she drives. With respect to ourselves, we develop habits of moving in the course of learning our bodies and learning to move ourselves, ways of doing that are at bottom coordinated dynamic patterns that in our adult life run off in consistent ways and that, being easily carried out and familiar, are commonly experienced at

---

[9]Sensations, however, may in some instances coalesce. A throbbing sensation, for example, may develop into a kinetic form. See Sheets-Johnstone 2006.

the margins of awareness: we tend not to be focally present to them. On the contrary, we are commonly focally attentive on accomplishing something, something as common and trivial as tying a shoelace or unlocking a door. Habitual movement patterns can, however, be made focally present. We can, for example, become aware of the dynamics of brushing our teeth. Indeed, were someone else to brush our teeth, we would immediately recognize that someone else was brushing our teeth, not just because we were not holding the tooth brush and not only because we could see someone in front of us holding and moving our toothbrush, but because *we would feel a foreign dynamics inside our mouth*. In short, when we turn attention to habitual movement patterns, we recognize our own kinetic melodies, indeed, our own *kinesthetic* melodies (Luria 1966, 1973); they bear the recognizable stamp of our own qualitatively felt movement patterns, our own familiar coordination dynamics.

Actually, to invoke 'embodiment' in some form in speaking or writing of action, experience, and so on, is adultist. Surely one would not describe an infant as an 'embodied agent' as it imitates an experimenter's tongue protrusions, for example, or as having an 'embodied experience' as it sucks on a nipple. Moreover if one asked an infant, "What matters most to your developing knowledge of the world?," he or she would answer "movement." As noted child psychologist Jerome Bruner observed, an infant's primary interest is in "agentivity," that is, agent and action (Bruner 1990). It is in and through movement that infants and young children discover aspects of themselves and of the world about them, aspects that do not disappear with age but that continue to inform their lives from beginning to end. Adult humans in various academic disciplines neglect movement and kinesthesia and indeed overlook having initially and originally learned their bodies and learned to move themselves. Their neglect and oversight are not rectified nor rectifiable by the term "embodiment" and its derivatives, and this because the coordination dynamics that develop in infancy and that perdure as foundational building blocks throughout our lives testify not to 'embodiment' but to animation, primal animation, and, in a complementary way to what prolific researcher and writer on coordination dynamics J.A. Scott Kelso aptly describes an "intrinsic dynamics", a dynamics grounded in the self-organizational patterns of living beings (Kelso 1995). Primal animation and an intrinsic dynamics infuse our being and define our aliveness; they are our point of departure for living in the world and making sense of it. An adultist stance overlooks these animate beginnings, these initial ventures into and explorations of movement. It overlooks as well the complex and subtle ways in which these literally animate beginnings were – and still are – integrally and inherently entwined with cognition and affect.

An investigation of our own habits teaches us about these animate beginnings; it teaches us about movement and kinesthesia directly, about how affects motivate and inform our movement, and about the built-in cognitive structures of movement. It teaches us how the particular coordination dynamics we articulate in walking, for example, are the result of the composite qualities of movement that we instantiate when we walk: it teaches us the basic fact that any movement creates its own space, time, and force, and thus a particular felt qualitative dynamic.

On the basis of this basic fact, it teaches us that we can change any habitual qualitative dynamic if we wish and instantiate a different dynamic. We can in fact make the familiar strange, and in so doing, discover what esteemed Russian physiologist Nicolas Bernstein termed 'degrees of freedom' in human movement. We can in other words create a range of different coordination dynamics, not only in changing our usual manner of walking – making it less tense, for example, or smoother, or more expansive – but in changing the way in which we write our name or move through any number of other everyday acts. We can furthermore make the familiar strange at the level of sheer movement itself by examining the degrees of freedom in turning to the side. For example, we might initiate the turn by a sideward extension of the leg, a twist at the waist, a twist at the shoulders, or a twist of the head, and the turn itself might be sharp, slow, slight, or sizable, its possible variations being virtually limitless. In sum, the "limitations" of voluntary movement are sizable freedoms, precisely as Bernstein showed. Because they are, learning our bodies and learning to move ourselves are not necessarily learnings restricted to infancy.

Evolutionary Biology and the Existential Fit of *Leib* and *Körper*

As has been emphasized, a sizable number of researchers consistently neglect movement and its experiential foundations in kinesthesia. The oversight is in truth surprising if not appalling in contexts where discussions of kinesthesia would be enlightening. Two examples from cognitive science readily make the point. In a near opening paragraph in a chapter on "Consciousness and Control of Action" in a section of *The Cambridge Handbook of Consciousness* titled "Cognitive Psychology," psychologist Carlos Umtilà states, "In the present chapter, I am concerned exclusively with motor (i.e., bodily) actions" (Umtilà 2007, p. 327); in an opening sentence in a chapter on "The Development of Consciousness" in a section of the same book titled "Developmental Psychology," psychologists Philip Zelazo, Helena Gao, and Rebecca Todd state, "This chapter examines the extent to which consciousness might develop during ontogeny" (Zelazo et al. 2007, p. 405). In the one instance, it is not only that without the parenthetical clarification there might indeed and for good reason be confusion about use of the term 'motor' with respect to animate actions, but that living consciousness is conjoined with a *motorology* in the first place and not with *kinesthesia*. In fact, no mention is made of kinesthesia at all. In the second instance, it is as if normal infants could be born and develop with no consciousness at all. As noted above, we come into the world moving; we are precisely not stillborn. Our animation is tightly bound to our tactile-kinesthetic bodies and to kinesthesia – in a word, to our consciously felt and consciously moving bodies. Moreover we are by nature *drawn to movement*, not only as Bruner's infant research shows, but as psychiatrist René Spitz's experimental studies of infants show. In fact, one might justly elaborate on Spitz's claim that we are by nature drawn to "the percept of the human face and eyes" (Spitz 1983, p. 149), namely, by noting that "the human face and eyes" *move* and consistently present

themselves to us *in movement*: eyes in looking and scanning, mouth in talking and singing, eyebrows in surprise and consternation, head in turning and tilting, the whole configuration of face in smiling and frowning, and so on.

Evolutionary biology and our own evolutionary heritage are relevant in just this context. To begin with, on the basis of his study of Hymenoptera, Darwin noted that "It is certain that there may be extraordinary mental activity with an extremely small absolute mass of nervous matter" (Darwin 1981, p. 145). He went on to observe specifically that "the wonderfully diversified instincts, mental powers, and affections of ants are generally known, yet their cerebral ganglia are not so large as the quarter of a small pin's head. Under this latter point of view, the brain of an ant is one of the most marvellous atoms of matter in the world, perhaps more marvellous than the brain of man" (ibid.). Clearly, members of the animal kingdom survive, if they do survive, not just because they are adept physically, but because they are consciously adept across a spectrum of faculties, precisely as Darwin indicates: they are affectively and cognitively attuned to their surrounds. They are animate in the full sense of being affectively and cognitively alive to themselves and to their surrounding world. Just such affectively and cognitively attuned living bodies are not distinct from, but of a piece with their physical bodies. They are thereby adept at making a living in a world that is never quite the same from 1 day to the next. Evolutionary anthropologist William Howells makes this point deftly and sharply in his observation that "hands and a big brain would not have made a fish human; they would only have made a fish impossible" (Howells 1959, p. 341).

To be consciously adept physically is to enjoy a livability in the world, to be existentially fit. An essential element of that fitness – that livability – is an awareness of one's own movement, without which instincts, mental powers, and affections, whether of ants or of any other creatures, would count for naught: an animal that knew not *how*, *what*, *that*, or *when* it was moving would be incapable of effective agency, indeed of agency at all.[10] The animal would not in fact be livable. Its livability, like its agency, is contingent on kinesthesia. The relationship between agency and kinesthesia is significant and is aptly exemplified in the well-known phenomenon of infant imitation. As elucidated at length elsewhere in a constructive phenomenology of animation (Sheets-Johnstone 1999, pp. 260–271), the capacity of infants to imitate mouth gestures (Meltzoff 1990; see also Gallagher and Meltzoff 1996; Meltzoff and M Keith Moore 1977, 1994), is tied not to a "body schema" or a "supramodal representational system" but to a tactile-kinesthetic body that is dynamically attuned to the world. Similarly, the "psychological primitive" that is there from the start (Meltzoff 1990) is not a crude stratum of intelligence but a burgeoning capacity to think in movement, a capacity that is foundational to adult human knowledge as well as a capacity that is clearly evident in animals such as tigers who hunt for a living. Misunderstandings and distortions of proprioception and kinesthesia occlude recognition of both a dynamically

---

[10]We might note that agency is empirically linked not to an ontological entity called a self but to an epistemological subject in the form of animation and kinesthesia.

attuned tactile-kinesthetic body and a capacity to think in movement. Misunderstandings and distortions in fact abound in discussions and explanations of infant imitation, often because, rather than hewing to empirical analyses tied to corporeal matters of fact, they are tethered to, or veer off into, theoretical entities, models, and exegeses. The error is not only on the side of cognitive scientists. An example from phenomenological writings just as readily shows how wayward theses and conclusions about infant imitation result from a neglect and even ignorance of movement and kinesthesia. The example is an fact chosen precisely because its focal concern is infant imitation.

In an article titled "Facial Embodiment in 'Invisible' Imitation," philosopher Beata Stawarska proposes and defends the thesis that an infant, in imitating, has a *visual* sense of its own face, which she calls "facial embodiment." This visual sense of its own face is a matter of the infant's "*detect[ing]*" that it is the target of the gaze of another in the process of imitating that other (Stawarska 2003, p. 149); that is, it "reads clues about the facial exterior proper [i.e., clues about its own face as a visible object] and develops the sense of being visible in a third-person mode" (ibid., p. 153). To support the credibility of its "read[ing] clues" and experiencing itself as visible "in a third-person mode" (ibid.), Stawarska "follow[s]" Baron-Cohen's postulation of an EDD (Baron-Cohen 1995), an "eye-direction detector," which she limns as "one of the mechanisms that allows the infant to recognize where the other is looking," adding that "This mechanism participates in establishing early self-other relations" (Stawarska 2003, p. 149). The fact that an EDD is a wholly hypothetical entity, a conjured "mechanism" in the brain, in essence a fantasized feature or "stall" along the cerebral mall, never surfaces. While Stawarska is at pains to show that infant imitation is not dependent on an internal representation any more than it is dependent on seeing oneself in a mirror or on some other form of specular representation (see Piaget 1962; Lacan 1977), she fails to understand the dynamic nature of both kinesthesia and proprioception, and indeed, to neglect kinesthesia near entirely, oddly mentioning at one point "internal kinesthetic sensations" (Stawarska 2003, p. 142), as if there were *external* kinesthetic sensations, and as if kinesthesia were a matter of sensations and not of an unfolding qualitatively felt kinetic dynamic in the first place, that is, the kinesthetic dynamic of protruding one's tongue. Moreover to speak on the one hand of young infants as having a "poor *motor* mastery of the body proper" (ibid., p. 145, italics added), of their learning gradually "to control and adjust their *motor* performance" (ibid., italics added), and of their "self's *motor* experience" (ibid., p. 148, italics added), and on the other hand, to speak of proprioception as "nonconscious, physiological information," (ibid., p. 146), and of "proprioceptive awareness" being "'a felt experience of the bodily position', such that one knows where a given bodily part is located without having to monitor it visually" (ibid., p. 147), and in fact to speak on both hands together of "exclusively motor non-perceptual proprioceptive information" (ibid.) is to elide any and all understanding of kinesthesia and developing coordination dynamics. The idea that "a sighted person's sense of facial expressions proper must *exceed* the sheer proprioceptive feedback" (ibid., p. 155) and the conclusion that in fact "facial embodiment proper exceeds proprioceptive feedback" (ibid., p. 158)

bypass not only phenomenological but "real-life, real-time" dynamic understandings of kinetic/kinesthetic/proprioceptive experience and even neurology.[11]

In bypassing recognition of the experienced dynamics of movement with a third-person orientation and vocabulary, Stawarska relies on "embodiment" to do the work of understanding both "the body proper" – *le corps propre*, the *Leib* – and its developmental relationship to *le corps tout court*, the *Körper* – the body "in a third-person mode." The experience of "being seen" – a "third-person" experience as Stawarska herself specifies – is, as indicated, an experience requiring a pointedly self-conscious awareness of oneself as *object* of another's gaze. Though Stawarska assures us that an infant "can feel herself to be the terminus of that gaze at a very early age" (ibid., p. 149), it surely stretches empirical credibility to think that newborns and young infants "at a very early age" (ibid.) who imitate the mouths gestures of others are self-reflective in this way. The marriage of *Leib* and *Körper* is in fact accomplished linguistically and mechanistically via "embodiment" instead of developmentally and experientially in a socially maturational sense. Indeed, what Hanne De Jaegher and Ezequiel Di Paoli term "participatory sense-making" in the context of offering an analysis of social cognition as an interactional dynamic process is nowhere in sight (De Jaegher and Di Paolo 2007); nowhere in sight either are intercorporeal sense-makings (Sheets-Johnstone 2008). The subtext thesis driving the enterprise is "*sensible reversibility*" (Stawarska 2003, p. 158), a vindication of Merleau-Ponty's notion of an "intertwining" or chiasm of touched and touching, seer and seen.

Infant imitation is exemplary of a host of topics in phenomenology and cognitive science that are in actuality grounded in the affective-cognitive-kinetic dynamics of animation. Shifting concern to this ground requires a shift in thinking – a paradigm shift in Thomas Kuhn's words. It requires thinking directly, intently, and unwaveringly along the lines of the body, reflecting at length on the aliveness of living bodies and all that that aliveness *by its very nature* encompasses in the way of movement, feeling, and cognition, all in an experiential sense. In imitating an adult's mouth gestures, a newborn infant is learning its body and learning to move itself. It engages in a "kinetic-kinesthetic *dynamic* matching" (Sheets-Johnstone 1999, p. 261), a transfer of sense from the visual body of another to its own tactile-kinesthetic body, discovering the dynamic possibilities and actualities of its own moving body in the process. Though not described or analyzed in such terms, infant imitation experiments testify to this fact just as they testify to responsivity and ratification of meaning (ibid., Consciousness). Such testimonials notwithstanding, an infant's actual transfer of sense from the visual body of another to its own tactile-kinesthetic body is unexplicated. While its impetus toward matching clearly lies in its own kinetic liveliness – its "primal animation" (Sheets-Johnstone 1999)

---

[11] Indeed, to speak of proprioception in terms of "proprioceptive feedback" is, to begin with, to speak in terms of a motorology, not of a living and lived-through dynamics as it unfolds and of that living and lived-through dynamics as a kinetic melody at the level of neurophysiological innervations and denervations (with respect to the latter, see Luria 1966, 1973).

and its "intrinsic dynamics" (Kelso and Scott 1995) – and in its dynamically attuned body, its actual transfer of sense remains topologically unexplained: how, in broad terms, does the infant know that the kinetic deformations it sees are replicable by kinetic deformations it can achieve?

The answer lies in both movement and topology, namely, in the fact that one changes shape as one moves, and in moving changes shape in invariant ways. The topological connection is experienced. In other words, self-movement is topologically distinct. It is experienced as distinct because topological specificity is inherent in the dynamics of self-movement, that is, in kinesthesia. Thus, when a fetus, even at 11 weeks, opens and closes its mouth, and later, at 4.5–5 months, slips its thumb into its mouth and sucks it (Furuhjelm et al. 1976), it is becoming topologically familiar with its body, specifically its mouth, the first and pivotally central topological attractor of human infancy (Spitz 1983; note too Furuhjelm et al. 1976, p. 52: "We develop from the head downward"). In short, the kinesthetic/kinetic dynamics of movement play out with topological specificity both on and within the body: tongue protrusions felt by the viewer and seen by the infant are in turn felt by the infant and seen by the viewer. Just as an infant's responsivity and its ratification of meaning have their origin in the *sensu communis* that is movement (Sheets-Johnstone 1999, Consciousness), so do its foundational topologically-informed transfers of sense. Indeed, its responsivity and ratification of meaning are grounded in topologically-informed transfers of sense rooted in the *sensu communis* of movement, movement that is both self-movement and the movement of others. (For more on intercorporeal sense-making, see Sheets-Johnstone 2008.)

In sum, pointed attention to lacunae in both present-day cognitive science and phenomenology shows that, whether a matter of science or philosophy, when a motorological or ontological "embodied" spin of one kind and another is put on animate sensibility, cognition, and affect, there is a distancing from animation and the experienced dynamics of the qualitative kinesthetic realities that inform the lives of animate forms. It is of interest to point out that the ground floor of cognition – animation – was recognized by Husserl not only in his abiding concept of *animate organism*, but in *his* concept of the phenomenon of "intertwining," which does not underwrite a reversibility, whether of the touched and the touching, the seer and the seen, or both, but a *unity* of mind and body – a "Bodily [i.e., *Körper*]-spiritual unity" (Husserl 1989, p. 352). Both concepts warrant careful and rigorous study. They validate in penetrating and acute ways the essentially dynamic nature of life itself and its "*double reality*" (ibid., p. 353).

References

Baron-Cohen S (1995) Mindblindness: an essay on autism and theory of mind. MIT Press, Cambridge, MA

Bruner J (1990) Acts of meaning. Harvard University Press, Cambridge, MA

Darwin C. 1981 [1871]. The descent of man and selection in relation to sex. Princeton University Press, Princeton, NJ

De Jaegher H, Di Paolo E (2007) Participatory sense-making: an enactive approach to social cognition. Phenomenol Cogn Sci 6(4) (forthcoming)

Dorsett DA (1976) The structure and function of proprioceptors in soft-bodied invertebrates. In: Mill PJ (ed) Structure and function of proprioceptors in the invertebrates. Chapman & Hall, London, pp 443–483

Furuhjelm M, Ingelman-Sundbert A, Wirsén C (1976) A child is born, revth edn. Delacourte Press, New York

Gallagher S, Meltzoff AN (1996) The earliest sense of self and others: Merleau-Ponty and recent developmental studies. Philos Psychol 9(2):211–233

Gallagher S, Varela FJ (2003) Redrawing the map and resetting the time: phenomenology and the cognitive sciences. In: Thompson E (ed) The problem of consciousness: new essays in phenomenological philosophy of mind. Calgary University Press, Calgary, Canada, pp 93–132

Gallese V (2001) The 'shared manifold' hypothesis: from mirror neurons to empathy. J Conscious Stud 8(5–7):33–50

Gallese V (2007) Intentional attunement. the mirror neuron system and its role in interpersonal relations. Interdisciplines. "What do mirror neurons mean?" http://www.interdisciplines.org/mirror/papers/1, Last accessed May 19, 2007

Gibbs R (2006) Embodiment and cognitive science. Cambridge University Press, New York

Hanna R, Thompson E (2003a) Neurophenomenology and the spontaneity of consciousness. In: Thompson E (ed) The problem of consciousness: new essays in the phenomenological philosophy of mind. University of Calgary Press, Calgary, Canada, pp 133–161

Hanna R, Thompson E (2003b) The mind-body-body problem. Theoria et Historia Scientiarum VII(1):23–42

Howells W (1959) Mankind in the making. Doubleday & Co., New York

Hurley SL (1998) Consciousness in action. Harvard University Press, Cambridge, MA

Johnson M (1987) The body in the mind. University of Chicago Press, Chicago, IL

Husserl E (1989) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: Book 2 (Ideas II) (trans: Rojcewicz R, Schuwer A). Kluwer, Boston, MA

Kelso, J. A. Scott (1995) Dynamic patterns: the self-organization of brain and behavior. Bradford Book/MIT Press, Cambridge, MA

Lacan J (1977) The mirror stage as formative of the function of the I as revealed in psychoanalytic experience. In: Sheridan A (trans) Écrits: a selection. W. W. Norton, New York

Lakoff G, Johnson M (1999) Philosophy in the flesh. Basic Books, New York

Laverack MS (1976) External proprioceptors. In: Mill PJ (ed) Structure and function of proprioceptors in the invertebrates. Chapman & Hall, London, pp 1–63

Lawrence DH (1932) Apocalypse. Viking Press, New York

Luria AR (1966) Human brain and psychological processes (trans: Haigh B). Harper & Row, New York

Luria AR (1973) The working brain (trans: Haigh B). Penguin Books: Harmondsworth, Middlesex, England

Lyon ML, Barbalet JM (1994) Society's body: emotion and the 'Somatization' of social theory. In: Csordas TJ (ed) Embodiment and experience: the existential ground of culture and self. Cambridge University Press, Cambridge, pp 48–66

Meltzoff AN (1990) Foundations for developing a concept of self: the role of imitation in relating self to other and the value of social mirroring, social modeling, and self practice in infancy. In: Cicchetti D, M Beeghly (eds) The self in transition: infancy to childhood. University of Chicago Press, Chicago, IL, pp 139–164

Meltzoff AN, M Keith Moore (1977) Imitation of facial and manual gestures by human neonates. Science 198:75–78

Meltzoff AN, Keith Moore M (1994) Imitation, memory, and the representation of persons. Infant Behav Dev 17:83–99

Merleau-Ponty M (1963). In praise of philosophy and other essays (trans: Wild J, James E). Northwestern University Press, Evanston, IL

Merleau-Ponty M (1968) The visible and the invisible (ed: Lefort C, trans: Lingis A). Northwestern University Press, Evanston, IL

Niedenthal PM, Barsalou LW, Winkielman P, Krauth-Gruber S, Ric F. (2005). Embodiment in attitudes, social perception, and emotion. Pers Soc Psychol Rev 9(3):184–211

Nöe A (2004) Action in perception. MIT Press, Cambridge, MA

Piaget J (1962) Play, dreams and imitation in childhood. W.W. Norton, New York

Petitot, Jean, Francisco J. Varela, Bernard Pachoud, Jean-Michel Roy (1999) Naturalizing phenomenology. Stanford, CA, Stanford University Press

Plato. Statesman. In: The dialogues of Plato, vol. 2 (trans: Lowett B). Random House, New York

Roy J-M, Petitot J, Pachoud B, Varela FJ (1999) Beyond the gap: an introduction to naturalizing phenomenology. In: Petitot J, Varela FJ, Pachoud B, Roy J-M (eds)Naturalizing phenomenology: issues in contemporary phenomenology and cognitive science. Stanford University Press, Stanford, CA, pp 1–80

Sheets-Johnstone M 1966 [1979, 1980]. The phenomenology of dance. University of Wisconsin Press, Madison, WI. Second editions: Dance Books, Ltd., London; Arno Press, New York

Sheets-Johnstone M (1986) Existential fit and evolutionary continuities. Synthese 66:219–248

Sheets-Johnstone M (1999) The primacy of movement. John Benjamins, Amsterdam/Philadelphia

Sheets-Johnstone M (2003) Kinesthetic memory. Theoria et Historia Scientiarum VII(1):69–92

Sheets-Johnstone M (2006) Essential clarifications of 'self-affection' and Husserl's 'sphere of ownness': first steps toward a pure phenomenology of (human) nature. Cont Philos Rev 39:361–391

Sheets-Johnstone M (2007) Schizophrenia and the 'Comet's tail of nature': a case study in phenomenology and human psycho-pathology. Philoctetes J (in conjunction with NY Psychoanalytic Institute) 1(2):5–45. This article is a target article with commentaries by Jaak Panksepp and Luc Ciompi, and Louis Sass, and a response by the author

Sheets-Johnstone M (2008) The roots of morality. Pennsylvania State University Press, University Park, PA

Sheets-Johnstone M (2009) Animation: an essential, fundamental, and properly descriptive concept. Continental Philos Rev (forthcoming)

Sheets-Johnstone M (2009) The corporeal turn: an interdisciplinary reader (Chapter 3). Imprint Academic, Exeter, UK

Sokolowski R (1972) Discussion. In: Tymieniecka A-T (ed) Analecta Husserliana, vol. 2: The later Husserl and the idea of phenomenology. D. Reidel Publishing, Dordrecht, pp 64–77

Spitz R (1983) Dialogues from infancy In: Emde RN (ed). International Universities Press, New York

Stawarska B (2003) Facial embodiment in 'invisible' imitation. Theoria et Historia Scientiarum VII(1):139–161

Thompson E (2007) Mind in life: biology, phenomenology, and the sciences of mind. Harvard University Press, Cambridge, MA

Toombs SK (2001) The role of empathy in clinical practice. J Conscious Stud 8(5–7):247–258

Umtilà C (2007) Consciousness and control of action. In: Zelazo PD, Moscovitch M, Thompson E (eds) The Cambridge handbook of consciousness. Cambridge University Press, Cambridge, pp 327–351

Varela FJ (1996) Neurophenomenology: a methodological remedy for the hard problem. J Conscious Stud 3(4):330–350

Varela FJ (1999) The specious present: a neurophenomenology of time consciousness. In: Petitot J, Varela FJ, Pachoud B, Roy J-M (eds) Naturalizing phenomenology: issues in contemporary phenomenology and cognitive science. Stanford University Press, Stanford, CA, pp 266–314

Wright BR (1976) Limb and wing receptors in insects, chelicerates and myriapods. In: Mill PJ (ed) Structure and function of proprioceptors in the invertebrates. Chapman & Hall, London, pp 323–386

Mill PJ (ed) (1976) Structure and function of proprioceptors in the invertebrates. Chapman & Hall, London

Zahavi D (2002) First-person thoughts and embodied self-awareness: some reflections on the relation between recent analytical philosophy and phenomenology. Phenomenol Cogn Sci 1:7–26

Zahavi D (2005) Subjectivity and selfhood. Bradford Book/MIT Press, Cambridge, MA

Zelazo P, Gao H, Todd R (2007) The development of consciousness. In: Zelazo PD, Moscovitch M, Thompson E (eds) The Cambridge handbook of consciousness. Cambridge University Press, Cambridge, pp 405–432

Empirical and Phenomenological Studies of Embodied Cognition

David Morris

The first grand wave of twentieth century cognitive science and philosophy of mind sought and still seeks to show how, contra Cartesian dualism, cognition is materialized in the brain. A new wave is rising from this, though, one that studies cognition not merely as "embrained" (Damasio 2000, 118; Collins 2000) but as embodied in a much stronger sense, as inseparable from and shaped by the concrete extra-cerebral structures and dynamics of the body, and the body's embeddedness in the natural and social world. This wave is gathering momentum. Whereas a decade ago researchers needed to protest that cognition is in fact strongly embodied, one can now find studies that leap right to the problem of how and in what sense it is so. This is a telltale sign of a nascent paradigm shift. Yet, in studying this new wave one finds that philosophy and science are still – for reasons recently remarked (see Dreyfus 2007; Gallagher 2007) and soon discussed – at sea for lack of conceptual frameworks to bring embodied cognition back to firm land. Brooks, whose robotics research is part of the new wave, puts it succinctly: "perhaps at this point we simply do not *get* it, and there is some fundamental change necessary in our thinking." (1991; cited in Dreyfus 2007, 251)

Since the study of embodied cognition surges with deeply conceptual problems and is in the midst of a sea change, this chapter cannot simply summarize fully worked out results. It is inevitably synthetic. Given the handbook's overarching goal, my aim is to show how an older wave of philosophy – the phenomenological tradition initiated by Husserl – can, on the conceptual level, complement new wave empirical results in cognitive science.

Phenomenology is well positioned for this complementary role. If philosophy in general is in the business of creating or analyzing the concepts we deploy in our everyday and scientific labours, phenomenology's distinctive task is to have rigorous description of the phenomena drive philosophy's conceptual endeavour. And whereas science normally fits empirical data into existing conceptual frameworks, shifting paradigms only in moments of crisis, phenomenology begins with a principled conceptual crisis – the reduction – that puts our most basic concepts into question.

D. Morris
Department of Philosophy, Concordia University, 1455 Maisonneuve Blvd.
West, PR-202, Montreal, Quebec, Canada H3G 1M8

The reduction is never left behind and it radically shifts paradigms all the way down. But it shifts paradigms by holding itself responsible to what empirically presents itself. So, as Russon (2006, 308) notes, phenomenology is a form of empiricism in which we "let our (rigorously enacted) observations specify the terms and parameters of our theories, rather than holding description answerable to theory." Phenomenological and empirical studies of embodied cognition can therefore mutually constrain or enlighten one another in ways that let phenomenologically inspired researchers contribute to new wave research (see, e.g., Varela 1996, 1999; Gallagher 1997, 2005; Thompson 2007; also see Schmicking in this volume, Petit 2003; Borrett et al. 2000). But, given the nature of the conceptual problem that is outlined below, instead of pursuing this strategy of mutual constraint, I lean more heavily on phenomenology's basic conceptual results, whilst keeping phenomenology in conversation with empirical studies.

The next section develops the conceptual problem by introducing and explaining the strong thesis that cognition is embodied and outlining the phenomenological view of it.

Empirical Studies of Embodied Cognition and the Spectres of Crypto-Cartesianism

What could it mean to challenge traditionally neurocentric views by making the strong claim that cognition is 'embodied over and above being embrained', by claiming, for example, that the hands have some role in cognition? And why would one ever say such a thing? One reason is to exorcise hidden, crypto-Cartesian[1] conceptual prejudices about body and mind that do not fit the empirical evidence and consequently hobble traditional approaches. As shown below, empirical studies pressure science into discarding traditional concepts of the body as inadequate in accounting *even for walking*, an obviously embodied activity. So we should be all the more suspicious of deploying such traditional concepts in trying to unify body and mind. But we will see that even as we try to escape these problematic concepts our tradition tangles us in further crypto-Cartesian prejudices – which is why we should turn to phenomenology as deriving results on a different conceptual level.

Empirical studies give evidence that human walking is controlled not by neurology alone but is shaped by leg-environment dynamics (Thelen 1984, 1995; Thelen and Smith 1994) and even social factors (cf. Zelazo 1983, 1984; Fogel 1993). Further, there is evidence that muscular synergies crucial to limb control are not specified by fixed anatomical structures but by "soft assembled," dynamic, environmentally modulated functions (Turvey 2007). Studies of biped, quadruped, and hexapod animals show that their gait pattern is best explained in terms of structured leg dynamics inherited through evolution, rather than central neural controllers

---

[1] The degree to which Descartes himself endorses the crypto-Cartesian and Cartesian prejudices here discussed is controversial. I use these terms broadly, to refer to positions derived from or referred to Descartes, whether rightly or wrongly.

(Full and Farley 2000). Robots that control walking through central processors cannot handle shifting terrain and require far more energy and actuator precision than are found in humans. These problems can be tempered by leveraging the dynamic morphology of leg joints, and of the body as linking legs together, for control. (Brooks 1991; Collins et al. 2005) Simulations show that body symmetry (vs. asymmetry) facilitates control of locomotion and increases locomotory efficiency (Bongard and Paul 2000).

While traditional accounts conceptualize limb control as a problem solved by the brain, the empirical studies show that *the dynamic body is itself* integral to control. Where traditional accounts thinly conceive the brain as a computational device (whether serial, parallel, or neural/distributed), the new wave thickens this by showing how the body itself is integral to the brain's controlling 'computations'. The mind not only extends into notepads, etc., that do cognitive work (as Clark and Chalmers (1998) argue), it extends into limbs that do work in their own control. The traditional, thin concept of the body as a dumb machine is thereby also thickened, because the body itself, in virtue of its dynamic morphology, has a kind of 'know how' or "cleverness" (Csepregi 2006).

Embodiment thus matters over and above embrainment – embodiment does cognitive work. To put it more sharply: The body is not a passive puppet moved by the brain as sole controlling agency. The body (with its brain) *is* a controlling agency.

Were we to extend this thesis to cognition in general we would say: the body is a *cognizing* agency. In this case the infamous "brain in a vat" (conjured to show, contra Cartesianism, how cognition can be embodied by way of being embrained) would be a misguided, vatic abstraction. More, it would *in fact be disembodied*, for it lacks the right kind of body to do thinking, namely the moving body so vital to kinetic coping and evolution. (See Sheets-Johnstone 1999b, 1990) But can we extend this thesis to cognition, which is after all, our topic? An objector will surely say that the above data, while convincing in showing embodiment's importance to actions such as walking, has no bearing on cognition proper.

In reply, we must first of all emphasize that walking is not merely limb movement, but intelligent interaction with the environment. So points about walking are already points about a discerning, judging, cognitive activity. But such activity has its roots in the body. For example, the complex locomotory response of cockroaches blasted with compressed air is too quick to be explained by nerves and is instead rooted in viscoelastic musculoskeletal structures (Dickinson et al. 2000, 103). The structure of the cricket body and ear, over and above neurology, is integral to its orienting to the chirps of other crickets. (Clark 2001, 127–128) Sheet-Johnstone (1999b; 2003b) shows in detail how bodily, animate movement already is intelligence. The body itself is a mindful responder. Perhaps neurology merely complicates the already supple adaptive richness of bodily responses (see Wheeler 2005, 134, 195–211).

Besides, the objection is premised on sharply dividing action, perceptual cognition and 'higher order' cognition from one another, so as to rule out drawing conclusions from one domain into the others. But thinkers as diverse as Dewey (1972), Merleau-Ponty (1962), Varela (1991, Varela et al. 1993), Hurley (1998), Berthoz (2000), O'Regan and Noë (O'Regan and Noë 2001a, 2001b; Noë 2004), Wheeler (2005), Thompson (2007),

and Turvey (2007, 663) argue that the traditional action-perception-cognition distinction is conceptually and/or empirically misguided. As Gallagher (2005, 136) insists, "[p]erceptual experience is generally accepted to be fundamental to other modes of cognition and action."

This point is amply confirmed by empirical evidence that action, perception, and embodiment mix in shaping 'higher' cognition. Just some of this evidence will now be mentioned. For example, psychologists show that walking is integral to the infant's cognitive development (Kermoian and Campos 1988; Fogel 1993; Thelen 2000). Tactile cognition is shaped by movements, skills and habits of touching (Katz 1989; Lederman and Klatzky 1987). Similarly, visual cognition is shaped by looking (Merleau-Ponty 1962; Churchland et al. 1994; O'Regan and Noë 2001a; Thompson 2007). Rhythms of the moving body influence our hearing metrically ambiguous six beat musical rhythms as being either in three groups of two beats, or two groups of three beats. (Philips-Silver and Trainor 2007)

On the level of language and thinking, studies show that sentence formation is not wholly governed by abstract syntax but is influenced by bodily speech rhythms alien to vatic embrainment (Lee and Gibbons 2007) and that body language is integral to thought and language, rather than being a merely external accompaniment (McNeill 1992). Gesture also facilitates human learning. Specifically, in learning, bodily gestures that are initially topologically isomorphic to a learning domain are metonymically shortened in ways that give rise to more abstract typological symbols that facilitate the learning of abstract concepts (e.g., gestures mimicking atomic attraction contract into shorthand symbols for atomic bonds) (Roth and Lawless 2002). Chimpanzee manipulation of what Clark (2006, 294) calls "material symbols" suggests that such symbols can facilitate high order cognition, for example, they facilitate chimpanzee judgement that the relations of difference and sameness are different relations. Together these points might suggest that the *gestural body itself* is a material symbol whose flexibility is integral to cognition – that we learn not by abstract symbol manipulation but by way of generating new symbols through our body's manipulation of things. This is a point well anticipated by Merleau-Ponty (1962) in his studies of expression and helpful in addressing the notorious symbol grounding problem (see, e.g., Thompson 1997).

It also resonates with Lakoff and Johnson's analyses of how our thinking and imaging is organized by schemata rooted in bodily ways of doing things. For example, the schema of something moving from a source, along a path, to a goal – a schema initially grasped on a bodily level – underlies metaphors in multiple cognitive domains (e.g., "I am trying to reach the end of this chapter"). (Lakoff and Johnson 1999; Johnson 1987, 2007; also see Gibbs and Berg 2000) Similar phenomena suggest that the roots of thinking (Sheets-Johnstone 1990) and even mathematics (Lakoff and Núñez 2000) are in the moving body. Data on mental imagery tasks (such as watching a moving object and reporting its final position from memory) further suggest that imaging involves a kind of internal gesture or material symbol manipulation rooted in our bodily relation to things (vs. an abstract symbol transformation) (Gibbs 2006; Gibbs and Berg 2000; Amorim and Isableu 2006). Researchers have also demonstrated what should probably be called the Proust

effect: recall of biographical episodes is faster when body position during recall is congruent (vs. incongruent) with body position during the original episode. (Djikstra et al. 2007; cf. Casey 2000; Sheets-Johnstone 2003a).

Over and above the brain, the morphology, dynamics and temporality of the body, and our evolved, moving attitude to the environment, shape and lurk in our acting, perceiving, speaking, learning, remembering, and conceptualizing. This link between the body and cognition is supported in recent neurological literature, which shows that 'higher' cognitive processes (involving activation of 'higher' brain structures) in fact also activate and are activated by 'lower' brain structures known to be active in emotional affect, in proprioreceptive and visceral processes, and in body control. As Tucker puts it, it "may be that there are no brain parts for cognition, at least not separate from the brain mechanisms pertaining to bodily functions." (Tucker 2007, 14; also see Damasio 1999, 2000, 2003; Freeman 1999; Gallagher 2005; Berthoz 2006; Thompson 2007) All of this, plus the observation that – as William James (1950, 138; cf. Sheets-Johnstone 1999b, ch. 10) notes – the brain evolved as an organ of action at one with the body, make it empirically implausible to think that there would be anything like *human* thinking in a vat other than our body, or that thinking *in general* is a purely embrained matter. Tragic evidence for these points is given by the closest thing we have to empirical studies of brains in vats, namely notorious studies at McGill university by Ewen Cameron and Donald Hebb (see Kubzansky 1961) of people deprived of sensory stimulation and self-movement. On the Cartesian or embrainment view, an immobilized mind would simply have nothing much to think about. But people who are deprived of 'embodiment over embrainment' lose their minds – their thinking is severely disturbed, implying that the mind is deeply mobile and bodily.

To review the above, by way of condensing and elaborating Wilson's (2002, 626) identification of six views of embodied cognition: (1) Cognition is situated in a real-world environment and inherently mixes perception and action. (2) Cognition is responsive to real-time demands and operates through the real-time dynamics and rhythms of the body (vs. being a rate-independent computational process). (3) The environment and the body do cognitive work, and (4) are part of cognition. (5) Cognition is for action. (6) Even 'off-line' activities such as remembering, imagining and abstract thinking have bodily roots. Cognition, in other words, is physically grounded (see Anderson 2007), in the sense that the physical specificities, rhythms, dynamics and shape of the moving body, and its embeddedness in the world and social settings, matter to cognition in a full-fledged way (vs. it being the case that only basic physics matters to cognition, in realizing a wetware computer or connectionist pattern recognizer).

New wave researchers endorse the above claims in various mixtures. The strong version of embodied cognition pursued here would endorse all of them. To briefly voice such claims through the work of important phenomenologically influenced researchers, Sheets-Johnstone (1999b) argues that taking evolution and the phenomena seriously means realizing that:

> At their most fundamental level, subjective experiences are tactile-kinesthetic experiences. They are experiences of one's own body and body movement; they are experiences of animate form. These experiences are the bedrock of thinking. (435)

As Gallagher (2005) puts it, the "body *actively organizes* its sense experience and its movement in relation to pragmatic concerns" (142) – the body is no mere passive receptor – and "a full picture of human cognition can be drawn only by exposing the details" of this (133).

Dreyfus (1992, 1998, 2002, 2007; Wrathall and Malpas 2000) approaches this issue by emphasizing that bodily, skillful coping is no mere accompaniment to cognition but is *itself* cognition. In Thompson's (2007, 256) words, "[s]ensory stimulation does not *cause* experience in us, which in turn *causes* our behaviour," rather (citing O'Regan and Noë) "skillful activity … *is* the experience." Noë (2004) details how bodily action is internal to perception and shapes consciousness, concluding that a plausible account of consciousness "must be an account of [it] as a natural phenomenon" (vs. an abstract computational phenomenon) and this "will be a tale, not about the brain, but about our active lives." (231) Thompson (2007, 128) strikes a deeper conceptual chord by arguing for an underlying "deep continuity of life and mind," such that "life and mind share a basic set of organizational properties, and organizational properties distinctive of mind are an enriched version of those fundamental to life." In other words, the phenomena drive these thinkers to conceive mind as a process rooted in evolution and development (both individual and social). Mind leverages what already belongs to life and to skillfully moving, organized bodies. As Johnson (2007, 279) writes:

> The human mind is not contained in the body, but emerges from and co-evolves with the body.…
> A human being is a body-mind, that is, an organic, continually developing process of events.

Why are these claims so new-wave, controversial or hard to grasp? Conceptual issues are unavoidable here, for *noticing the empirical phenomena catalogued above is not yet the same as having these phenomena count as evidence for strong (or even weak) versions of the embodied cognition thesis*. One can easily conceptualize the phenomena so as to stick with old-wave embrained views. For example, evidence shows that emotional affect is integral to our ability to make cognitive judgements (e.g., to decide which of two projects to pursue). This is what leads Damasio (1999, 2000, 2003) to urge 'embodiment over embrainment'. But as Gallagher (2005, 135) points out, in actual detail Damasio deflates the body to its representation in the brain (or inflates it to ideas of the body). For Damasio, the body is important – *but only as embrained*. Data of embodiment is muddled into data of embrainment without even noticing it. Likely we can always (whether deliberately or unthinkingly) muddle the data this way. For example, physically grounded body rhythms can be muddled into relations between neural time-stamps registering bodily events. What tells us to stick with the strong embodiment thesis and what helps us pursue it without muddling cannot be mere data *but conceptual frameworks and desiderata*.

But our navigation of this conceptual level is swamped by crypto-Cartesian prejudices that run so deep as to distort our account of the body, never mind mind–which is why I began with studies of walking. These prejudices keep rising from the crypts of our tradition, like conceptual vampires that suck the life from cognition. Even as we nail Cartesian substance dualism into the coffin of dead ideas, it shape-shifts into a new dualism of brain and body. Even as we exorcise the Cartesian concept of mind, we

retain an essentially Cartesian concept of the body as dumb machine (see Merleau-Ponty 1962, 1965; Burwood et al. 1999), and thence a dualism between the body as living agency and the body as passive machine (see Thompson 2007, 230–237; Leder 1992) – all of which betray bodily phenomena. Indeed, Wheeler (2005, ch. 2–3) catalogues *eight* crypto-Cartesian prejudices, tracing their cross-linked origins in Cartesianism and their persistence in philosophy and science. Most are converse to the claims of embodied cognition treated above, but it is worth noting the prejudices of explaining cognition in terms of (1) subject-object dualisms, (2) representations (that repair the subject-object dualism) and their transformations, (3) temporally austere terms (such that, e.g., physically grounded real-time rhythms muddle into rate-independent relations between representations); (4) and explaining intelligence as the outcome of general purpose reasoning (vs. reasoning specific to a creature's evolution).

To pursue the embodied cognition thesis we must shed these crypto-Cartesian conceptual habits and see the phenomena anew. But here a deeply subtle point comes into play. Old-wave research conceptualizes embodiment as a solution to the problem of how mind fits with matter. Conceptually, the problem is how the phenomenon of mind, *which can be coherently described in terms independent of the body*, is materialized. It is answered by appeal to properties of the body, *which (correlatively) can be coherently described as independent of mind*. Old-wave philosophy and science accept this framing of the problem but endeavour to resolve it by showing how the descriptive differences can be bridged. Crucially, *this formulation of the problem remains essentially Cartesian and dualist in character*. It is no wonder that efforts to resolve it lapse into crypto-Cartesian prejudices or retain residual dualisms. (The embrainment thesis in effect acknowledges that if body and mind are descriptively dualised, then the solution requires a special body part (the brain as, e.g., symbol processor) capable of doing something most unbodylike.)

Phenomenology radically reframes the problem, indeed it shows that the old problem is badly put. To anticipate, phenomenological description shows that the mind is an inherently temporal process, that cognitive contents are not given all at once, but take time. But the mind does not supply or regulate this time. The mind is rather its exposure to a time that exceeds it. Husserl finds that this exposure has the form of kinaestheses, felt bodily movements.[2] This is why Thompson can insist that skillful activity is not the cause of experience, but "*is* the experience":

---

[2]The deployment of kinaesthesis at this point raises the problem that the concept of kinaesthesis might already depend on a distinction or indistinction between mind and body, such that the overall argumentative strategy is troubled. (My thanks to Daniel Schmicking for pointing out this problem.) We can venture a reply that, descriptively, mind and body appear as *relative* terms within the more primordial phenomenon of kinaesthesis. We do not find notes of a melody outside of a melody, and vice versa. Similarly, we do not find coherent moments of kinaesthesis outside of their overarching temporal organization, and vice versa. The thought here is that the different temporal moments through which a kinaesthesis unfolds stand as the relative body of this kinaesthesis, and the organizing, temporalizing flow internal to and overarching these moments stands as the relative mindfulness of the kinaesthesis. The moments and their temporalizing, the bodilyness and the mindfulness, are not two different things, although they are distinguishable. In this view, mindfulness is exposure to time via the bodilyness of movement, but mindfulness is this exposure as a temporalizing that runs through the moments. Pursuit of this point would, though, take another paper.

the 'stuff' of experience is time – and time is felt in movement. Along these lines, the strategy of the next section is to show that mind *is* body by showing that the time of thinking *is* the time of bodily moving. The section after that conversely suggests that body *is* mind by showing that the time of bodily moving is not the time of a clockwork machine, but of cognizing. We typically fail to grasp these points and find the strong embodied cognition thesis controversial because (as the next section suggests) we prejudicially remove body and mind from time, reducing them to things completely present and given, e.g., to Cartesian substances or computational or bio-mechanical systems. And we also prejudicially tend to reduce time itself to an already given dimension. By drawing on the theme of time, the following sections are meant to give an example of the sorts of arguments, conceptual frameworks and resources that phenomenology can offer to complement recent empirical results on embodiment. They also give an example of phenomenology's strategy and approach to mind and body.

Mind Is Body: Movement, Time and the Prejudice of Presence

I develop a phenomenological argument that mind *is* body by linking Husserl's *Cartesian Meditations* (CM, 1991) with his lesser known emphasis on kinaesthesis.[3] *CM* begins by endorsing Descartes's argument that philosophy, for methodological reasons, must begin with the "I think." But Husserl argues that Descartes muddles his beginning by prejudging what the "I think" is. So Husserl sets out to rigorously describe the phenomenon of thinking. He ultimately finds that the "I think" is already a kinaesthetic "I can." In other words, phenomena of mind, *rigorously described in their own terms*, must *already* be conceptualized as bodily phenomena.

How does Husserl arrive at this point and what does it mean? One of Husserl's fundamental insights is that cognitive phenomena always have a *horizonal* character (see esp. *CM* §19). Phenomenological description shows that in perceptual experience cognitive objects (*noemata*) are not given all at once. To perceive a die is not to have it presented or represented as an entirety in a cognitive instant. Each actually appearing die-face has its sense as a real die-face only through its relation to other actually appearing faces and through their portending of hidden faces, it has its sense in the way that one die face leads to another. Hidden faces are only potentially apparent, but in turning the die they can become actually apparent. The sense of the real die as a whole is thus not something entirely present, but an invariant pattern of an (in principle) unending unfolding of actual and potential perceptual profiles of the die. The structure of ""predelineated" potentialities" through which such a thing unfolds is what Husserl calls a horizon. (*CM* 82) Horizons are

---

[3] On this topic and its neglect in the reception of Husserl, see Sheets-Johnstone (1999b), Zahavi (1994), Zahavi (2003).

dynamic (the die's predelineated potentialities change as one turns it) and complexly nested (the sense of the die's six-face lies in its dots as horizonally portending each other).

The above point is about *noematic horizons* – the horizons of *objects* of cognition. But Husserl shows that each *act* of cognition (*noesis*) also has an horizon – a *noetic horizon*. The cognitive act of perceiving this-here real die has its sense in further acts that ongoingly fulfill the initial act's sense (the sense "die" is in the act of seeing the six-face turning into the act of seeing the two-face). So a cognitive act "is never present to actual consciousness as a finished datum; it becomes "clarified" only through explication of the given horizon and the new horizons continuously awakened." (*CM* 82–83) Both noetic acts and noematic objects take time.

Already the above means that perceptual cognition is of itself bodily. First, as Thompson (2007, 248) observes, following Husserl and Merleau-Ponty (1962, 1964b), the fact that perceptual consciousness always unfolds in correlation with unfolding object profiles means that objects and cognitive acts *themselves* mark perception as endogenously spatial and perspectival – which "requires that we be embodied," have a bodily place from which we perceive. Second, as Landgrebe (1973) emphasizes, according to Husserl, each experience, as having its sense in a horizonal "indication of possibilities," has the form of: "I can continue and will presumably gain this or that new impression from the thing." (10) That is, experience is not just perspectival: it involves spatial perspectives *changing over time*. Moreover, horizonal possibilities of change are not represented in some sort of completed roster, rather I am aware "of them as that which is not yet actual but that can take place precisely because *I can* bring it about." (10, my emphasis) But according to Husserl, the "I can" here is *bodily*: the protentive promises of the die's faces, in which rests the very sense "die," is confirmed or disconfirmed in *kinaestheses*, movements initiated and felt in experiencing the die. (8) So "[p]erception is impossible without the experience we gain in kinaesthetic movement." (11) The phenomenological evidence shows that the Cartesian "I think," as taking time, must already be preceded by an "I can" and an "I move." The fifth meditation of *CM* emphasizes that this in virtue of what Husserl conceptualizes as *Leib*, the lived body (also translated as "animate organism"). The *Leib* is the body felt and lived in the "I can move," a body that is inalienably my own. *Leib* contrasts with *Körper*, a term designating the body as an object studied, say, by the doctor.[4] The "I think" is in and of itself *already* bodily, for thinking is determinate only by way of temporally opening to further possibilities that exceed presence – and this is by way of the moving *Leib*. (See also Husserl (1989, §§59–60), Sheets-Johnstone (1999a,b, 131–140, c), Zahavi (1994, 2003, 98–109), cf. Gallagher (1986b)).

It is worth reflecting here on the root of Husserl's divergence from Descartes. In the second meditation Descartes notoriously argues that I see the wax not with my eyes but with my *mind*. Extraordinarily, Descartes's argument, which concludes by splitting the "I think" from the bodily "I can," is driven by *the very same horizonal*

---

[4] On *Leib* vs. *Körper* see Leder (1992).

phenomena that drive Husserl to the opposite conclusion. For Descartes, the wax is so excessively horizontal that not even the imagination could exhaustively determine its identity. This requires an "idea," which alone can determinately comprehend all variations of the wax for all time. Seeing the wax is therefore an act of judgement that subsumes phenomena (that are in principle indeterminate) to an idea (that is in principle fully determinate). The body could never pull off such an act because its very time of being is an opening to always excessive indeterminacies – bodily time inherently invites doubt. So if a Cartesian idea is to escape indeterminacy and doubt, it must clinch its object via a representational relation. This representational relation is highly problematic. By virtue of the very problem an idea is meant to solve, an idea must have a fully determinate content that subsumes or intends something indeterminate. So an idea must subsume something entirely unlike an idea. Dualism is thus foreshadowed in the very notion of an idea.

What is going on here? Faced with the phenomenon of horizons as threatening the foundations of determinate knowledge, Descartes reaches to ideas as something that would be fully determinate in face of indeterminacy. But the phenomena give no evidence of fully determinate entities. Objects are certainly not given to us as fully determinate. If they were, there would be no doubt and no need for philosophy. And the phenomena show Husserl that cognitive acts are also not wholly determinate but horizonal. Likely, Descartes bypasses the indeterminacy of the phenomena because of a prejudice that runs deep in all our philosophy and thinking: the prejudice of presence, which motivates his appeal to a cognitive content – an idea – that would be entirely present, given and determinate all at once, that would subsume the indeterminacy of changing time to the timeless determinacy of unchanging essences.

For Husserl, Descartes's problem is the solution: the very indeterminateness of the horizon, which yet has "a determinate structure" (*CM* 83) that cannot, however, be reduced to presence, is what carries the sense that Cartesian ideas would determine as all present. Brooks (1991) insightfully moots the problem that robots run into when we engineer them to navigate by fully representing the determinancies of an environment that is in fact inexhaustible and ever changing: he suggests that the environment be used as is its own 'representation'. Husserl similarly moots the Cartesian problem that the phenomena are inherently indeterminate and thence exceed us: he notes that the phenomena in their very temporal indeterminacy are the very determinacy needed for knowledge. (As Husserl puts it, "The [horizonal] predelineation itself, to be sure, is at all times imperfect; yet, with its *indeterminateness*, it has a determinate structure." (*CM* 83)) Indeed, if phenomena were wholly present and determinate, then truth and verification would have no sense. As Merleau-Ponty (1962, 204) would note, the six faces of a real die cannot all be present to us in the same way at once; the reality of die lies in the way it keeps on temporally turning up new faces and hiding other ones from us.

In other words, on its deepest register, all the crypto-Cartesian prejudices discussed above in turn presume the *prejudice of presence*: the prejudice that being and beings, whether thingly or cognitive, are, ontologically speaking, given all at once as fully there. To undo this prejudice we must, like Brooks and others discussed

above, conceive cognitive processes as assembling themselves in real-time by way of being bodily processes movingly embedded in worlds and times that exceed them. But this conceptual move is easy to muddle, and we can now say why: to stick with it we have to overcome the prejudice of presence and our related tendency to think that explanation is satisfactory only if it comes to a stop with something that is all there at once (which also means that we tend to explain temporal processes by making time into something that is all there).

For example, Lloyd (2004; cf. Varela 1999), who is inspired by Husserl's emphasis on cognition's inherent temporality, notes our tendency to render cognitive states as being entirely present in the brain areas that 'light up' in fMRI imagery. He also notes that we forget that fMRI averages brain activity over time such that it blurs changes on the rapid time scale of thinking. In reducing cognition to fully determinate activations we prejudicially neglect the way cognition might actually be a matter of rapid micro-temporal rhythms of change, kin to horizons of neural activity enacted in concert with bodily horizons. In this context the objection that kinaestheses and *Leib* can (for the brain in the vat) be simulated mistakenly reduces these to presence by abstracting them from the real-time demands that drive our brainy-body evolution, in the way that real rhythm is ruined by reducing it to ratios. But our technology and conceptual frameworks likewise tend to mistakenly envat the real brain in a frozen present. Notably, the embodied cognition thesis can always be refuted by way of reducing its terms to presence. Ultimately the only thing that testifies against this reduction is the being of the phenomenon itself in its own time. But that is precisely what the prejudice of presence freezes out of the picture.

Here we must address another objection. The above may convincingly show that perceptual cognition, perhaps even imagining and dreaming, are already bodily. Surely, though, this analysis does not go through with *abstract* cognition! But let us look at a doctor. As her patient walks through the door, her head cocks to listen for shuffling feet that betray neurological problems, her nostrils sniff for breath odours symptomatic of metabolic problems; later she taps tympanically on the patient's torso, listening for inner organ problems. The doctor experiences her diagnostic reasoning unfolding *in these kinaestheses*. If medical reasoning could be abstracted from bodily kinaestheses, then likely it would not be fraught with the stupendous errors and prejudices that Groopman (2007) catalogues. Later, when writing up her report, she feels her reasoning operating in the hesitation of her pen and body, which hesitation *is* her way of thinking that: "But maybe this logical option is not the case." (See Johnson 2007, ch. 3) Scientific reasoning, too, proceeds through kinaesthetic engagement with images and apparatus (see Galison 1997), and the process of writing things up and conversing about them.

Our earlier point that noetic activity is horizonal indicates a deeper phenomeno-logic behind these last points. If my thought "triangle" did not or could not actually temporally lead through its horizon to the thought that its angles sum to 180° and so on (and to noematic objects verifying these), then I would either doubt that I am really thinking "triangle," or think that I am changing its ideal sense (by thinking, say, of spherical triangles). The sense of even ideal objects lies in the fulfilment of their horizon potentials. But for me as human, not divine, fulfilling these horizons

takes bodily doing and time. If you put me into a sensory deprivation tank or a 1984-like polity that perverts all recording and communication, then I shall go mad, for I shall lose track of the sense of my own thinking. As Husserl argues in his later *Crisis of the European Sciences* (Husserl 1970) and "The Origin of Geometry" (available in Merleau-Ponty 2002), community, history, "writing down," and moving one's *Leib* in a shared life-world are crucial to abstract ideas and thinking, a point amplified in the *cogito* chapter of Merleau-Ponty's *Phenomenology of Perception* (1962; also see Merleau-Ponty 2002). From this perspective, the Cartesian "evil genius," who could even make us doubt that 2 + 2 = 4, is not an all powerful mind but … disembodiment itself. To put the brain in any 'vat' other than the body is to subject thinking to such ineradicable doubt that it is driven mad or mistakes itself for an all knowing divine. In contrast, the mind conceptualized as *Leib* is itself open to endless temporal horizons that at once account for doubt (without which we could not be philosophers) and for its resolution via kinaesthetic, intersubjective explorations and recordings (without which we could not be scientists).

As Husserl compactly puts it, in one of the most pungent sentences of *CM*: "Objects exist for me, and are for me what they are, only as objects of actual and possible consciousness." (99) But, as Husserl immediately insists, we need to give a concrete account of this possibility and actuality. In doing so we find that the kinaestheses of the *Leib*, as the kernel site of actuality and possibility, are indispensable to *all* objects – including abstract ones – such that, as Gallagher (2005, 1–5, 133–138) might put it, prenoetic bodily processes are prelude and postlude to each noesis. The mind is not so much given in the head, yet subsequently extended into things, as Clark and Chalmers (1998) have it. The mind *is* its extension in living moving bodies, its opening in time.

Aside from his discovery of horizons, one of Husserl's other great discoveries is intentionality: that all consciousness is consciousness *of* something, that consciousness is of itself a relation to a content. For the tradition we are criticizing, the phenomenon of intentionality is a great problem that repeats the Cartesian problem of ideas: how can a subjective representation be intentionally related to an outside object wholly unlike it? What we are in effect learning is that intentionality is in fact the solution to the Cartesian problem – once we understand that intentionality is *bodily and inherently temporal*, rather than being the presence of a (fully present) cognitive state to an external object.

On this analysis, the phenomenological labours of Heidegger and Merleau-Ponty are (merely) crucial inflections of Husserl. Merleau-Ponty (1962) deepens the point that intentionality is inherently bodily, through his well known discussion of the intentional arc, *Leib* and bodily habits and movement. (See Barbaras 1992, 2003; Gallagher 1995, 1986a, 1986c; Gallagher and Marcel 1999; Dreyfus 2002; Leder, 1990, 1992) Heidegger's *Being and Time* (Heidegger 1962) expands Husserl's treatment of the moving body by showing, in effect, that the sphere of the moving *Leib* is in fact the complete sphere of interpretation that is our intersubjective world: the prenoetic body is in fact prenoetic human reality (cf. Dreyfus, 1991, 1992; Dreyfus and Hall 1982; Olafson 1987). And *Being and Time* further shows

how the proper conception of human reality entails overturning the prejudice of presence and temporal austerity. It requires thinking of being as time – and not reducing time to a present dimension.

The conceptual points articulated here by way of phenomenology complement the ones articulated earlier via empirical study. But phenomenology presents a strategic and methodological advantage in drawing its point directly from the phenomenon of thinking and flagging deep conceptual prejudices that we will have to undo if we are not to muddle our study of embodied cognition.

Body Is Mind: Bringing the Zombie to Leib

Above we saw that the phenomenon of mind, when rigorously described, is not something wholly present to itself but is an open temporal transcendence that exceeds itself. This is to say that mind *is* body, for the term "body" simply designates and describes the phenomenon in which this temporally open self-transcendence is figured and takes place (cf. Russon 2006). And so we can better grasp the claim that mind *is* body by better grasping how the body, as this temporally open self-transcendence, is thus already mind. I do this by briefly indicating three phenomenological arguments.

First, I pursue a point on the cusp of our intertwined claims that mind *is* body and body *is* mind. Once we abandon the prejudice of presence and turn to a temporally rich phenomenology of mind, we find (as Husserl showed in detail) that memory is crucial to mind, both as primary memory (which implicitly retains the unfolding present so that, e.g., we can hear a note as following the previous one) and as secondary memory (which explicitly recalls episodes of the distant past). But as Casey (2000), Sheets-Johnston (2003) and Russon (2003) observe, memory is not strictly speaking 'in the head': it is in our body and the places we inhabit. The body in its kinaesthetic engagement with places and markers is a well of memory integral to mind – the body in the world is itself the first "writing down." Correlatively this suggests that one's body already works as memorial mind.

Second, the thought that the human body is merely a dumb, mindless Cartesian machine is just the thought that the concept of a "zombie" is coherent. In philosophy of mind a zombie is (the equally fictional and unempirical) counterpart of the brain-in-the-vat: a body without a mind for a mind without a body, a creature who looks and acts just like you and me but is not in fact thinking or experiencing. Thompson (2007, 232) subjects this concept to a devastating critique via Husserl. Basically, if perception and perceptual behaviour have horizonal structures – and they must, given the phenomenon of perception – they entail experience of bodily kinaestheses, that is, awareness and anticipation of one's bodily kinaestheses. So if zombies are behaviourally indistinguishable from humans perceiving things, they must be experiencing, kinaesthetic bodies. So the zombie-body must be a *Leib*, not a *Körper*. The zombie who behaviourally treats the die as real must have a feel for the die turning in hand. I might add that if mind is descriptively characterized as

unfolding according to kinaesthetic horizonal structures, if mind is not some kind of presence over and above this, then a zombie-body moving according to mindful horizons and temporality is in its very movement already mindful. There is no further sort of thing for a mind to be, no further 'place' in which to find mind – although there are further developmental layers to bodily mindfulness, and these would be crucial to differences between mindfulness in the broad sense exhibited by animal bodies in general, and its elaboration, through the incorporation of language, tools, and culture, in the phenomena we find, for example, in human minds. The above suggests how movement of the *Leib* as whole is of itself mindful. This does not amount to behaviourism, for behaviourism claims that mind is nothing other than behaviour conceptualized *as something all present*. What the above asks us to see is rather that we must not conceptualize behaviour of the *Leib* as mere changes between states that are all present, but as ontologically escaping presence, via the temporality of behaviour.

Third, from Husserl onward, phenomenologists have long criticized the "theory theory of other minds" (TTM). TTM addresses the Cartesian problem of other minds. If there is nothing mindful about the body itself, and we do not encounter other minds directly but only through the bodies associated with them, how can we know that others around us have minds and are not merely zombies? TTM's answer is this: to explain the behaviour of other bodies, we find we need a theory, one that invokes another mind as a theoretical-explanatory term. We thus encounter other minds through theoretical inferences. Husserl, Merleau-Ponty (1962, 1964a), and other phenomenologists argue that we do not have the data necessary for making this sort of inference, and that in any case the phenomena testify against TTM. One encounters another not by way of theory, but by a more fundamental, prenoetic "pairing" between one's *Leib* and another *Leib*. (*CM* §§49-54) This counter-claim gains empirical support from developmental psychology and neurology. For example, the discovery of mirror neurons (which activate either when one makes a particular meaningful gesture, say tearing, or sees another doing it) suggests that one relates to others as meaningfully moving bodies who are counterpart to oneself (see Gallagher 2004). But this would mean that the other's body already appears as itself being a phenomenon of mind. And this would make a great deal of evolutionary sense. What we moving animals most need to do is look out for other moving animals as out to get or help us. From an evolutionary and phenomenological perspective, bodies that move in organized ways already appear to us as mindful. We Cartesians can doubt this, but we are sure not to when a zombie-body starts hunting us down with exceptional stealth and cunning.

Conclusion

Altogether, the phenomena and the imperatives of life testify that cognition is embodied in the strong sense. But we continually do not notice this, or we deflect telltale phenomena, by stepping outside the flow of time and life onto a temporally

austere platform through which we hope to secure the fully present terms that would satisfy certain prejudicial presumptions of explanation. Both empirical and phenomenological studies, however, suggest that this step is misguided. We must bring mind back to life, conceive body and mind as two aspects of the same continually developing temporal process.

References

Amorim M-A, Isableu B (2006) Embodied spatial transformations: "body analogy" for the mental rotation of objects. J Exp Psychol Gen 135:327–347

Anderson M (2007) Review of how the body shapes the mind. Philos Psychol 20:127–142

Barbaras R (1992) Motricité et phénoménalité chez le dernier Merleau-Ponty. In: Richir M, Tassin E (eds) Merleau-Ponty: phénoménologie et expériences. Jérôme Millon, Grenoble

Barbaras R (2003) Life and perceptual intentionality. Res Phenomenol 33:157–166

Berthoz A (2000) The brain's sense of movement. Harvard University Press, Cambridge, MA

Berthoz A (2006) Emotion and reason: the cognitive science of decision making (trans: Giselle W). Oxford University Press, Oxford

Bongard JC, Paul C (2000) Investigating morphological symmetry and locomotive efficiency using virtual embodied evolution. In: Meyer J-A, Berthoz A, Floreano D, Roitblat H, Wilson SW (eds) From animals to animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior. MIT Press, Cambridge, MA, pp 420–429

Borrett D, Kelly S, Kwan H (2000) Phenomenology, dynamical neural networks and brain function. Philos Psychol 13(2):213–228

Brooks R (1991) Intelligence without representation. Artif Int 47:139–159

Burwood S, Gilbert P, Lennon K (1999) Philosophy of mind. McGill-Queen's University Press, Montreal

Casey ES (2000) Remembering: a phenomenological study, 2nd edn. Indiana University Press, Bloomington, IN

Churchland PS, Ramachandran VS, Sejnowski TJ (1994) A critique of pure vision. In: Koch C, Davis JL (eds) Large scale neuronal theories of the brain: computational neuroscience. MIT Press, Cambridge, MA, pp 23–60

Clark A (2001) Reasons, robots, and the extended mind. Mind Lang 16(2):121–145

Clark A (2006) Material symbols. Philos Psychol 19:291–307

Clark A, Chalmers DJ (1998) The extended mind. Analysis 58:7–19

Collins HM (2000) Four kinds of knowledge, two (or maybe three) kinds of embodiment, and the question of artificial intelligence. In: Wrathall M, Malpas J (eds) Heidegger, coping, and cognitive science: essays in honor of Hubert L. Dreyfus. MIT Press, Cambridge, MA, pp 179–195

Collins S, Ruina A, Tedrake R, Wisse M (2005) Efficient bipedal robots based on passive-dynamic walkers. Science 307:1082–1085

Csepregi G (2006) The clever body. The University of Calgary Press, Calgary

Damasio AR (1999) The feeling of what happens: body and emotion in the making of consciousness. Harcourt Brace, New York

Damasio AR (2000) Descartes' error: emotion, reason, and the human brain. Quill, New York

Damasio AR (2003) Looking for spinoza: joy, sorrow, and the feeling brain. Harcourt, Orlando, FL

Dewey J (1972) The reflex arc concept in psychology. In: Axetell GE, Boydston, JA (ed) The early works, 1882–1898. Southern Illinois University Press, Carbondale, pp 96–109

Dickinson MH, Farley CT, Full RJ, Koehl MAR, Kram R, Lehman S (2000) How animals move: an integrative view. Science 288:100–106

Djikstra K, Michael PK, Rolf AZ (2007) Body posture facilitates retrieval of autobiographical memories. Cognition 102:139–149

Dreyfus HL (1991) Being-in-the-world: a commentary on Heidegger's being and time, Division I. MIT Press, Cambridge, MA

Dreyfus HL (1992) What computers still can't do: a critique of artificial reason. MIT Press, Cambridge, MA

Dreyfus HL (1998) The current relevance of Merleau-Ponty's phenomenology of embodiment http://ejap.louisiana.edu/EJAP/1996.spring/dreyfus.1996.spring.abs.html

Dreyfus HL (2002) Intelligence without representation – Merleau-Ponty's critique of mental representation: the relevance of phenomenology to scientific explanation. Phenomenol Cogn Sci 1:367–383

Dreyfus HL (2007) Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. Philos Psychol 20(2):247–268

Dreyfus HL, Hall H (1982) Husserl, intentionality, and cognitive science. MIT Press/Bradford Books, Cambridge, MA

Fogel A (1993) Developing through relationships: origins of communication, self, and culture. Harvester, New York

Freeman WJ (1999) How brains make up their mind. Weidenfeld & Nicolson, London

Full RJ, Farley CT (2000) Musculoskeletal dynamics in rhythmic systems: a comparative approach to legged locomotion. In: Winters JM, Crago PE (eds) Biomechanics and neural control of posture and movement. Springer Verlag, New York, pp 192–205

Galison P (1997) Image and logic: a material culture of microphysics. University of Chicago Press, Chicago, IL

Gallagher S (1986a) Body image and body schema: a conceptual clarification. J Mind Behav 7:541–554

Gallagher S (1986b) Hyletic experience and the lived body. Husserl Stud 3:131–166

Gallagher S (1986c) Lived body and environment. Res Phenomenol 16:139–170

Gallagher S (1995) Body schema and intentionality. In: Bermúdez JL, Marcel A, Eilan N (eds) The body and the self. MIT Press, Cambridge, MA, pp 225–244

Gallagher S (1997) Mutual enlightenment: recent phenomenology in cognitive science. J Conscious Stud 4(3):195–214

Gallagher S (2004) Understanding interpersonal problems in autism: interaction theory as an alternative to theory of mind. Philos Psychiat Psychol 11:199–217

Gallagher S (2005) How the body shapes the mind. Oxford University Press, Oxford

Gallagher S (2007) Review of Mark Rowlands, body language: representation in action. In: Notre Dame philosophical reviews http://ndpr.nd.edu/; the review is dated 2007.09.19, and is at http://ndpr.nd.edu/review.cfm?id=11183

Gallagher S, Marcel AJ (1999) The self in contextualized action. In: Gallagher S, Shear J (eds) Models of the self. Imprint Academic, Thorverton, UK, pp 273–299

Gibbs RW (2006) Embodiment and cognitive science. Cambridge University Press, Cambridge

Gibbs RW, Berg EA (2000) Mental imagery and embodied activity. J Ment Imagery 26:1–30

Groopman JE (2007) How doctors think. Houghton Mifflin, Boston, MA

Heidegger M (1962) Being and time (trans: Macquarrie J, Robinson E). Harper & Row, New York. Original edition 1927

Hurley SL (1998) Consciousness in action. Harvard University Press, Cambridge, MA

Husserl E (1970) The crisis of the European sciences and transcendental phenomenology (trans:Carr D). Northwestern University Press, Evanston, IL

Husserl E (1989) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. Second book: Studies in the phenomenology of constitution (trans: Rojcewicz R, Schuwer A) Kluwer, Dordrecht

Husserl E (1991) Cartesian meditations (trans: Cairns D). Kluwer, Dordrecht. Original edition 1931

James W (1950) The principles of psychology, vol 1. Dover, New York. Original edition 1890

Johnson M (1987) The body in the mind: the bodily basis of meaning, imagination, and reason. The University of Chicago Press, Chicago, IL

Johnson M (2007) The meaning of the body. The University of Chicago Press, Chicago, IL

Katz D (1989) The world of touch. Lawrence Erlbaum, Hillsdale, NJ

Kermoian R, Campos JJ (1988) Locomotor experience: a facilitator of spatial and cognitive development. Child Dev 59:908–917

Kubzansky PE (1961) The effects of reduced environmental stimulation on human behavior: a review. In: Biderman A, Zimmer H (eds) The manipulation of human behavior. Wiley, New York

Lakoff G, Johnson M (1999) Philosophy in the flesh: the embodied mind and its challenge to western thought. Basic Books, New York

Lakoff G, Núñez RE (2000) Where mathematics comes from: how the embodied mind brings mathematics into being. Basic Books, New York

Landgrebe L (1973) The phenomenological concept of experience. Philos Phenomenol Res 34(1):1–13

Leder D (1990) The absent body. University of Chicago Press, Chicago, IL

Leder D (1992) A tale of two bodies: the cartesian corpse and the lived body. In: Leder D (ed) The body in medical thought and practice. Kluwer, Dordrecht

Lederman SJ, Klatzky RL (1987) Hand movements: a window into haptic object recognition. Cogn Psychol 19:54–64

Lee M-W, Gibbons J (2007) Rhythmic alternation and the optional complementiser in english: new evidence of phonological influence on grammatical encoding. Cognition 105:446–456

Lloyd DE (2004) Radiant cool: a novel theory of consciousness. MIT Press, Cambridge, MA

McNeill D (1992) Hand and mind: what gestures reveal about thought. The University of Chicago Press, Chicago, IL

Merleau-Ponty M (1962) Phenomenology of perception (trans: Colin S). Humanities, Atlantic Highlands, New Jersey. Original edition 1945

Merleau-Ponty M (1964a) The child's relation with others. In: Edie JM (ed) The primacy of perception. Northwestern University Press, Evanston, IL, pp 96–155

Merleau-Ponty M (1964b) The primacy of perception and its philosophical consequences. In The primacy of perception. Northwestern University Press, Evanston, IL, pp 12–42

Merleau-Ponty M (1965) The structure of behaviour (trans: Fisher AL). Methuen, London. Original edition 1941

Merleau-Ponty M (2002) Husserl at the limits of phenomenology. Northwestern University Press, Evanston, IL

Noë A (2004) Action in perception, representation and mind. MIT Press, Cambridge, MA

O'Regan JK, Noë A (2001a) A sensorimotor account of vision and visual consciousness. Behav Brain Sci 24:939–1031

O'Regan JK, Noë A (2001b) What it is like to see: a sensorimotor theory of perceptual experience. Synthese 129:79–103

Olafson FA (1987) Heidegger and the philosophy of mind. Yale University Press, New Haven, CT

Petit J-L (2003) On the relation between recent neurobiological data on perception (and action) and the Husserlian theory of constitution. Phenomenol Cogn Sci 2:281–298

Philips-Silver J, Trainor LJ (2007) Hearing what the body feels: auditory encoding of rhythmic movement. Cognition 105:533–546

Roth W-M, Lawless D (2002) How does the body get into the mind? Hum Stud 25:333–358

Russon J (2003) Human experience: philosophy, neurosis, and the elements of everyday life. State University of New York Press, Albany

Russon J (2006) On human identity: the intersubjective path from body to mind. Dialogue: Can Philos Rev XLV:306–314

Sheets-Johnstone M (1990) The roots of thinking. Temple University Press, Philadelphia

Sheets-Johnstone M (1999a) Phenomenology and agency: methodological and theoretical issues in Strawson's 'The self'. In: Gallagher S, Shear J (eds) Models of the self. Imprint Academic, Thorverton, UK, pp 231–252

Sheets-Johnstone M (1999b) The primacy of movement. John Benjamins Publishing Company, Amsterdam

Sheets-Johnstone M (1999c) Re-thinking Husserl's fifth meditation. Philos Today SPEP Suppl:99–106

Sheets-Johnstone M (2003a) Kinesthetic memory. Theoria et Historia Scientiarum: Int J Interdis Stud 7:69–92

Sheets-Johnstone M (2003b) Movement: the generative source of spatial perception and cognition. Francine Dolins and Robert Mitchell, (eds.). Cambridge University Press

Thelen E (1984) Learning to walk: ecological demands and phylogenetic constraints. Adv Infancy Res 3:213–250

Thelen E (1995) Motor development: a new synthesis. Am Psychol 50(2):79–95

Thelen E (2000) Grounded in the world: developmental origins of the embodied mind. Infancy 1(1):3–28

Thelen E, Smith LB (1994) A dynamic systems approach to the development of cognition and action. In: Palmer S (ed) MIT Press/Bradford Books series in cognitive psychology. MIT Press/Bradford Books, Cambridge, MA

Thompson E (1997) Symbol grounding: a bridge from artificial life to artificial intelligence. Brain Cogn 34:48–71

Thompson E (2007) Mind in life: biology, phenomenology, and the sciences of mind. Belknap Press of Harvard University Press, Cambridge, MA

Tucker DM (2007) Mind from body: experience from neural structure. Oxford University Press, Oxford

Turvey MT (2007) Action and perception at the level of synergies. Hum Mov Sci 26:657–697

Varela FJ (1991) Organism: a meshwork of selfless selves. In: Tauber AI (ed) Organism and the origins of self. Kluwer, Dordrecht, The Netherlands, pp 79–107

Varela FJ (1996) Neurophenomenology: a methodological remedy for the hard problem. J Conscious Stud 3(4):330–349

Varela FJ (1999) The specious present: a neurophenomenology of time consciousness. In: Petitot J, Varela FJ, Pachoud B, Roy J-M (eds) Naturalizing phenomenology: issues in contemporary phenomenology and cognitive science. Stanford University Press, Stanford, pp 266–314

Varela FJ, Thompson E, Rosch E (1993) The embodied mind: cognitive science and human experience. MIT Press, Cambridge, MA

Wheeler M (2005) Reconstructing the cognitive world: the next step. MIT Press, Cambridge, MA

Wilson M (2002) Six views of embodied cognition. Psychon Bull Rev 9(4):625–636

Wrathall M, Malpas J (eds) (2000) Heidegger, coping, and cognitive science. MIT Press, Cambridge, MA

Zahavi D (1994) Husserl's phenomenology of the body. Etudes Phénoménologique 19:63–84

Zahavi D (2003) Husserl's phenomenology. Stanford University Press, Stanford, CA

Zelazo PR (1983) The development of walking: new findings and old assumptions. J Motor Behav 15:99–137

Zelazo PR (1984) Learning to walk: recognition of higher order influences. Adv Infancy Res 3:251–260

Part IV
Intersubjectivity

The Problem of Other Minds

Søren Overgaard

Introduction

The problem of other minds is a problem of modern philosophy that can be traced back to the French seventeenth century philosopher René Descartes. Descartes held the dualistic view that, apart from God, there are two kinds of things (or "substances") in the world: "extended things" and "thinking things". He thought, in particular, that a human being is a union of an extended thing (a body) and a thinking thing (a mind). He recognized that, given such a view, an account has to be given of how these two fundamentally different things are united, and in particular, how there can be causal interaction between them. But Descartes does not seem to have fully realized that another sort of issue needs to be addressed as well. Consider the following passage from Descartes' *Meditations on First Philosophy*:

> if I look out the window and see men crossing the square, as I just happen to have done, I normally say that I see the men themselves […]. Yet do I see more than hats and coats which would conceal automatons? I *judge* that they are men. (Descartes 1984, 21)

Descartes treats this as a subtle example of how we sometimes think we see something, have it displayed in front of our eyes, while in fact we simply judge or assume that it is there, without actually perceiving it. We thought we saw real human beings, but in fact we merely saw hats and coats that could as well conceal automatons. If we wanted to make sure, however, it seems we could lift those hats and coats and check whether we have judged correctly. But given Descartes' dualism, it is not easy to see how this move could be of any use. The trouble is that, precisely when you hold that the body is one sort of thing, and the mind a completely different sort of thing, and we only have a human being when the two are united in a particular way, then we can lift all the hats and coats we like without ever being able to ascertain that we are dealing with a human being. For what we will be able to see are merely the extended bodies: not the unextended, immaterial mind itself.

S. Overgaard
Center for Subjectivity Research, University of Copenhagen, Njalsgade 140-142, DK-2300 Copenhagen S, Denmark; and Philosophy Department, University of Hull, Hull, HU6 7RX, United Kingdom

A problem emerges, then. How is it possible to have knowledge of, or even reasonably justified beliefs about, the contents of other minds? If all I ever see is a twisting body, then how can I know that pain is being felt in that body? In fact, the problem goes deeper. For if I only have cognitive access to the external body, then how can I even have reason to believe that there is any mind there? And if there is no mind there, there is no real human being there. Descartes himself was not aware of this problem, nor were most philosophers in the centuries following Descartes (cf. Avramides 2001, 50 and passim). But in the twentieth century, the problem of other minds moved to the top of the philosophical agenda. Its repercussions are felt in many discussions within philosophy of mind and cognitive science today.

In this chapter I will suggest that the traditional responses to the problem of other minds are inadequate because they offer an insufficiently radical revision of the Cartesian conception of the mind. I will also argue that phenomenology offers the radical revision that is needed. In contrast to a variety of other approaches, the phenomenological analyses of embodiment and expression provide an account of the mental that makes social cognition intelligible.

The Reality of the Problem of Other Minds

There is a tendency, and not only among people who are not professional philosophers, to regard the problem of other minds as an idle philosophical conjuring trick that is irrelevant to the real issues of mind and cognition, and hence may simply be ignored. When the problem of other minds is introduced via Cartesian metaphysics, for example, as I have just done, many feel that it must be one that we can legitimately dismiss on the ground that nobody is a Cartesian dualist anymore. Philosophers have also frequently presented the problem as a *skeptical* challenge. We are supposed to come up with an account of our knowledge of the mental states of others that provides for at least some cases in which the possibility that we might be wrong is ruled out. If we cannot do that – if, even in the cases we single out as paradigmatic examples of knowledge of another's mental state, there will always remain the possibility, say, that the other is pretending – it is held that we have not solved the problem of other minds. Many have felt that whatever pleasure philosophers might derive from playing such games, they are surely of marginal importance. After all, at least when we are not wearing our philosophical hats, none of us considers it remotely possible that when our friends and relatives fall and twist their ankles, say (henceforth screaming and shouting), they really feel nothing at all.

However, it is a mistake to think the problem of other minds can be dealt with so swiftly. Whatever importance we might be inclined to attach to skeptical challenges, the other minds problem in its most interesting and fundamental form is not a skeptical problem. And it is not one that applies only to full-blooded Cartesians either. Rather, the problem of other minds sets up *a requirement that must be met by any satisfactory account of the mental*: it *must not be such as to make a mystery of social cognition*. Or, more positively expressed, any satisfactory account of the mental

must be able to make it intelligible that one mind can have knowledge of another mind. It is the problem in this form that I shall be concerned with in this article.

Conservative Responses to the Problem

It is useful to make a very rough distinction between two types of approach to the problem of other minds. One type of approach retains a crucial element – but not necessarily all the elements – of Descartes' position: the view, namely, that the mental life of one person cannot be directly perceived or experienced by another person. In other words, this type of approach assumes some kind of distinction between the body, and its behavior, gestures, and sounds, and the mental phenomena themselves. But it need not be committed, for example, to Descartes' view that the mental states are themselves the states of some immaterial, non-extended thing; it may remain noncommittal on this question, or even contradict the Cartesian view. All it is committed to is the view that one person cannot directly perceive another person's mental states. I will call this type of approach a *conservative* approach. The contrasting type of approach, unsurprisingly, is defined by its insistence on the claim that the mental *can* be directly perceived. It thus rejects the Cartesian element that the conservative position preserves. To stay with the political categories, I label this type of approach a *revisionist* one.

Conservative approaches have not always been the dominant trend. But it seems safe to say that they are today, both in philosophy, narrowly defined, and in the cognitive sciences. I will therefore discuss them in a bit more detail than the traditional revisionist alternatives.

The classic conservative approach is the *argument from analogy*. Probably the first clear articulation of the argument is found in John Stuart Mill's *An Examination of Sir William Hamilton's Philosophy* (cf. Mill 1979 (1865), 204–206). We may formulate the basic idea of the argument in the following way. Suppose I observe that whenever I get stung by a bee, take a fall, and similar things – that is, whenever my body is affected in certain ways by other bodies or objects – there is a characteristic sensation, pain, which tends to give rise to a characteristic pattern of bodily activity on my part, such as screaming, rubbing the place that was stung or hit, and so on. On the basis of such observations I conclude that whenever a B_1 event of a particular type occurs (a bee stings my arm), a corresponding M type of event occurs (I feel pain), which in turn gives rise to particular types of B_2 events (screaming, rubbing the arm, etc.). Now I observe other bodies, very similar to mine, moving about in ways that are again very similar to the ways in which I move my body around, and I notice that sometimes the types of B_1 event that tend to give me pain happen to other bodies as well. Moreover, I observe that generally, when such events occur, B_2 events tend to follow that are of the same types as those, which a sensation of pain would give rise to in my own case. Now, is it not reasonable for me to infer that the only missing link – an M type of event, a sensation of pain – is there when these things happen to other bodies, even if I do not observe it?

Whether or not this is how we actually form beliefs about other people's mental lives, the argument from analogy could still offer a rational *justification* for such beliefs. But, although there are exceptions (e.g. Hyslop 1995), most contemporary philosophers think the argument fails to do that. It is often observed that the argument is a case of *induction based on one case only* (cf. Carruthers 1991, 13). In one sense, of course, this is not true of the argument as I have just sketched it. For it may be based on thousands of observations of correlations between bodily input, mental state, and bodily output, in one's own case, and a similar number of observations of bodily inputs and outputs in the case of other people. But the problem is that the mental states as such only ever occur in my observations of myself. It is only in my own case that I can observe the mental link between input and output; and I can have no good reason for assuming that similar inputs and outputs could not be linked in entirely different ways in other bodies.

There is, however, a very different type of conservative account that may look more promising. After all, a major problem with the argument from analogy is its heavy reliance on my intimate knowledge of *my own* mental life, which does not seem to license inferences to other cases. What we might term the *best explanation argument* avoids precisely that (cf. Dennett 1971; Churchland 1988, 70–72). Let us assume, again, that the mental states of others really are unobservable. This in itself does not establish that we cannot have good justifications for believing that they exist. Nobody has ever seen an electron, but that hardly means that we have no reasons to believe electrons exist. On the contrary, electrons are entities that figure in a comprehensive theory of particle physics for which we have excellent scientific evidence. In general, we can say that "a theory about unobservables can be beliefworthy if it allows us to explain and to predict some domain of observable phenomena better than any competing theory" (Churchland 1988, 71). This is the case with the physical theory about electrons. Could it not be the case that the unobservable mental states of others – indeed, perhaps even one's own mental states – are similarly theoretical entities, belief in which is justified by their position in a theoretical framework that allows us to explain and predict observable bodily behavior? In other words, I do not infer that other people have feelings, sensations, etc., on the basis of an observed analogy between their behavior and my own behavior together with my intimate knowledge of how my own mental life is connected with my behavior. Rather, I infer that *we all* have feelings, sensations, thoughts and the rest because, even though none of these things can be perceived, it is a theory in precisely such terms that best explains and predicts the behavior that we all display. This theory that is supposed to embed our beliefs about (other) minds is usually labeled "folk psychology" (Churchland 1988, 71).

This account certainly does not suffer from the weakness that we found in the argument from analogy. But can it give us what we want? There is reason to think that it cannot. The best explanation argument requires us to stake our conviction that any of us have mental lives at all, on the ability of some "folk-psychological" theory to explain and predict behavior. All our talk about minds and mental states is now to be treated as plausible only to the extent that people's behavior "is best

explained and predicted in terms of desires, beliefs, perceptions, emotions, and so on" (ibid., 71). That is to say, if some other theory should become available, which made no references to such mental phenomena, then we should be prepared to abandon "folk psychology", including all beliefs involving desires, perceptions, and emotions. Indeed, that a superior theory already is available, one that explains and predicts behavior in the terms of neuroscience, is precisely the conclusion that prominent defenders of the best explanation argument want us to reach. But, quite apart from the other problems that such an "eliminative" view of the mental involves, it can hardly count as a satisfactory answer to the problem of other minds. The problem of other minds as I have formulated it assumes that others have mental lives that we can somehow have knowledge about, and challenges us to come up with an account of the mind that makes such cognition intelligible. To deny that there are any minds is thus to eliminate the phenomenon that needs explaining. And an explanation that only works if we commit ourselves to regarding minds and mental states as unobservable *and in principle dispensable* entities postulated by some theory designed to explain and predict behavior, is hardly an attractive one.

Conservative views are widespread in contemporary philosophy of mind and cognitive science. Conservatism is perfectly compatible not only with the view that the mental may ultimately be eliminated in favor of neuronal states and processes, but also with the view that the mental can be *reduced* to such states and processes. Those who reject the Cartesian view that each of us is an immaterial mind to be distinguished from the extended and observable human body in favor of the view that each of us is a "brain in a skull" (cf., e.g. Searle 1983, 230), are, from the point of view of the other minds problem, as conservative as Descartes himself. The "Cartesian materialist" who identifies the mind with the brain, and mental states with brain states, steers clear of invoking mysterious otherworldly entities, but follows Descartes in driving a wedge between observable, bodily behavior and the mental life itself. So in terms of the problem of other minds, the views of classical reductive materialists tend to be conservative, rather than revisionist.

In the contemporary debates concerning social cognition, most positions on offer may be classified as conservative. Both so-called "theory-theorists" and "simulation-theorists" generally assume that the mental phenomena of others are unobservable. The two accounts differ mainly in the explanations they offer of how we go about ascribing such unobservable phenomena to others. If we simplify matters slightly, we can say that most advocates of the "theory–theory of mind" endorse something like the best explanation argument, whereas many simulation theorists would advocate something that shares obvious affinities with the argument from analogy.

The main difference between prominent versions of theory–theory and the best explanation argument is that, whereas the latter can be construed as an account of our justification for attributing mental states to ourselves and others, which need not be committed to any account of how we actually go about making these attributions, theory–theory is intended precisely to offer the latter sort of account. Theory-theorists typically believe that since "the mental states of others

(and indeed of ourselves) are completely hidden from the senses, they can only ever be inferred" (Leslie 1987, 139). We infer and attribute these "unobservable states" by deploying a "theory of mind" (ibid.). One prominent theory-theorist, Alison Gopnik, argues that our ordinary psychological beliefs literally constitute a theory that we construct as children in order to explain our own behavior as well as the behavior of others (Gopnik 1993).[1] Intentional states and other mental states are thus "theoretical constructs" that may be revised in the light of evidence, and it is ultimately "an open question" whether they will have to "go the way of 'phlogiston'" (ibid., 12). The considerations advanced above against the argument from best explanation apply to this version of theory–theory as well. In addition, one may feel that to view young children as little scientists testing various theories in order to explain the behavior of bodies around them is to paint an overly intellectualistic picture of children's social lives (cf. Hobson 1991).

Simulation theorists avoid this difficulty by attributing to children and adults what Robert Gordon calls a "hot methodology" instead of the "cold", intellectualist methodology envisaged by theory–theory. A "hot" methodology is one "which exploits one's own motivational and emotional resources and one's own capacities for practical reasoning" (Gordon 1996, 11). One prominent version of the view has in common with the argument from analogy a reliance on one's intimate familiarity with one's own mental life (Goldman 2000). On this view, we attribute mental states to others by projecting ourselves into their shoes, as it were, and creating "pretend states" in ourselves intended to correspond to the states of the others. This account seems not to address the other minds problem as such. After all, what we want made intelligible is our conviction that there are any foreign "mental shoes" to project ourselves into. Goldman suggests a simulationist reply to this more fundamental problem as well, however. He directs attention to so-called mirror neurons (cf. Gallese and Goldman 1998) and suggests that these are but one subclass of a range of "resonance phenomena", some of which Goldman takes to be conscious. The idea, then, is that a human being may notice the same sort of experience within herself when she watches another creature engaged in some activity, as she experiences when she herself is engaged in the activity in question, and on that basis conclude that there must be similar experiences going one in that other creature (Goldman 2000). Obviously, however, this takes us back to a version of the argument from analogy, and an objection is immediately forthcoming: if a creature has the very same "inner experience" when it watches another body engaged in an activity as it has when it is itself engaged in that sort of activity, how could it ever occur to it to conclude that there was *another* subject of experience in that body? Surely, the sensible thing to conclude would be that those other bodies, too, must have some special connection with the creature's own mental life.[2]

---

[1] Other advocates of theory-theory favor an account in terms of innate theory-of-mind "modules" (for discussion, see Gopnik 1996).

[2] For a more detailed discussion of theory of mind, see Gallagher and Zahavi 2007, chapter 9.

Reductive Responses to the Problem

In part of the twentieth century, philosophers and psychologists favored a different approach to the problem of other minds. It was believed that one could reject the Cartesian element preserved by the conservative approach by reducing the mental to something observable. I will briefly discuss two such reductive approaches. *Behaviorism* in its crudest form maintains that mental phenomena really are behavioral phenomena. On this view, to be in pain consists in displaying a certain type of behavior (screaming, crying, etc.), or at least to be disposed to this kind of behavior. What Descartes thought were phenomena belonging to an immaterial mind distinct from the extended physical body are thus really phenomena ultimately to be explained in terms of the movements of that body, the noises it produces, etc. Behaviorism has had many adherents among psychologists and philosophers, and behaviorist ideas influenced much thinking about the mind in the early and middle decades of the twentieth century (e.g. Ryle 2000 (originally published 1949)). Some behaviorists believed that mental phenomena would ultimately be explained in terms of, indeed perhaps reduced to, "mere stimuli and mere movement", "colorless movement" (cf. Hull 1943, 26, 25).

Another once popular revisionist view is what is sometimes called *neutral monism*. The reason for mentioning this strange view here is that it has been endorsed by some of the major figures of twentieth century philosophy. Neutral monists argue that what Descartes supposed were two kinds of substance – immaterial minds and material bodies – are in reality merely two types of "logical construction" out of the same "neutral" stuff, namely "sense contents" or "sense data" (cf. Ayer 1990 [1936], 130; cf. Russell 1949, 36). Since other minds, on this view, are to be explained in terms of "the occurrence in my sense-history of the appropriate series of sense-contents" (Ayer 1990, 139), it follows that I have as much empirical evidence for their existence as I have for the existence of material objects (or even my own mind). The neutral monist approach, we should note in passing, was not entirely independent of the behaviorist one. An important first step on the way to reducing other minds to sense-contents in my sense-history was to reduce them to mere behavior, mere perceivable bodily movements and noises. The latter could then be explained in terms of sense-contents. Thus, it is no coincidence that neutral monists often espoused robust behaviorist views: unless other minds could be defined completely "in terms of the behaviour of their bodies" (Ayer 1990, 139), the prospects for the final reduction to sense-contents would be thwarted.

For various reasons, neither of these positions is on the philosophical and scientific agendas anymore. All we need to say about them as responses to the other minds problems is that they both seem to explain *away* that which we want explained. If crude behaviorism were true – if all there is is mere colorless bodily movement – there would not be any mental phenomena for us to attempt to make intelligible. If neutral monism were true – if other minds reduce themselves to series of sense-contents in *my* sense-history – then solipsism would follow.

It thus seems that neither the typical conservative responses nor the classic revisionist alternatives constitute very promising approaches to the problem of

other minds. If we abstract from all the details of the various positions and arguments, and try to focus on the fundamental thrust of both types of approaches, we might be able to understand why. The classic conservative approach retains the Cartesian gap between observable, physical behavior and unobservable (whether physical or non-physical) minds. And it seems as if this gap, once it is allowed to open up, is simply too wide for any inferences to bridge. The reductive revisionist reply, on the other hand, tries to close the gap by showing that what the conservatives suppose belongs to some independent mental side, really dissolves into elements of the physical side of the gap.[3] And this simply seems to do away with the phenomena we want explained. (As we have seen, the best explanation variant of the conservative approach threatens to give the same result.) So if none of these responses seems to work, what other ways may we attempt to address the problem of other minds?

Phenomenological Responses to the Problem

A useful point of departure for understanding the phenomenologists' take on the problem of other minds is the observation that there is a sense in which the revisionist responses that we have considered do not take their revision of the Cartesian view far enough (the last step of the neutral monist account excepted). For the aim of the behaviorist reduction that they attempt is to show that the phenomena that Cartesians and other conservatives think are essentially different from, and somehow hidden behind, mere physical movements and noises, really are nothing but such movements and noises, or can at least be adequately accounted for in terms of the latter. But that means they accept one half of the Cartesian picture – the picture of the body as a mere *res extensa* – and simply erase or ignore the other half.[4] This, all the major phenomenologists would insist, is not a sufficiently radical revision. For it leaves the reductive revisionists in agreement with the conservatives on the following, crucial point: *all we ever really see are the properties of a mere physical object* – the body. In contrast to this, all the phenomenologists attempt to articulate what might be called a *non-reductive revisionist* account; and they do so primarily, though not exclusively, by attempting a radical reinterpretation of the body.

However, before I offer a sketch of the main features of the phenomenological approach to other minds, I need to say a few words about what I understand by

---

[3] The neutral monists' (very implausible) further reductive step may be ignored here.

[4] In a recent paper on the history of psychology, Alan Costall argues that both the "introspectionism" usually believed to have preceded behaviorism as the dominant trend in psychology and Watsonian behaviorism "were arguing from the same premise, the *antithesis* of behaviour, on the one hand, and mind and consciousness on the other. Both were committed to an overly subjectivized conception of subjectivity, and an overly objectivized conception of behaviour" (Costall 2006, 649). Interestingly, Costall also argues that the current "cognitivist" paradigm, contra its own self-image, has not abandoned this Cartesian picture. The theory-of-mind debate would seem to support this claim (cf. Leudar and Costall 2004).

"phenomenology". I use "phenomenology" to refer to a way of doing philosophy that has been explicitly endorsed by figures such as Edmund Husserl, Max Scheler, Martin Heidegger, Jean-Paul Sartre, and Maurice Merleau-Ponty, and which has inspired psychologists, psychiatrists, sociologists, and others to employ similar methods in their work. Yet I do not want to reserve the label for thinkers who explicitly associate themselves with this brand of continental philosophy. As I understand "phenomenology" here, Ludwig Wittgenstein, for example, could be said to offer a phenomenological response to the other minds problem (cf. Overgaard and Zahavi 2009). The crucial feature of phenomenology, on the interpretation adopted here, is the attempt to let "phenomena" – e.g., human beings, trees, houses, gas bills, and general elections, taken just the way they present themselves to us or are "experienced" by us – guide our philosophical or scientific thinking. Thus, a phenomenologist is one who argues that when we think about what a mind is, for example, an important role is to be given to how we actually experience our own minds and mental states, and how we experience (if indeed we do) the minds and mental phenomena of others.

The phenomenological revision may be subdivided in a *direct* and an *indirect* response to the other minds problem. The indirect response is one that need not be explicitly intended as a reply to the problem of other minds. It consists of attempts to show phenomenologically that we cannot intelligibly ascribe certain types of mental phenomena to disembodied minds or brains in vats. Husserl (1989) and Merleau-Ponty (1962), for example, have argued extensively that at least the following two, closely connected conditions must be met for a subject to be able to perceive three-dimensional objects in space. First, the perceiving subject must itself be present in three-dimensional space. This point follows from phenomenological considerations of the way in which a perceived object presents itself in a certain perspective, as located in a certain direction and at a certain distance relative to the perceiver. Such perspectival appearances point back to the perceiving subject as the "center" or "zero point of orientation" (Husserl 1989, 166; cf. Sartre 1989, 317). Second, the perceiving subject must have an ability to move itself around in space, to effect changes in perspective and distance. Without a close connection between the "kinesthetic" activity of the perceiver and the flow of perceptual appearances, those appearances, so Husserl argues, cannot function as appearances of stable three-dimensional objects. Both points suggest that the perceiving subject must be an embodied or bodily subject. They thus attack the conservative view of the mind from the inside, as it were. They attempt to show that I cannot be a thinking (or perceiving) thing, unless I am also, in some sense, a bodily thing, and not just in the sense in which a brain is a type of "bodily thing". For without an ability to walk, run, swim, crawl, or in some other way actively explore my environment, so the argument goes, I would not have a perceptual environment – at least not one that resembles my actual perceptual environment.[5]

---

[5] For a contemporary version of the argument, see Noë 2004.

Now if I must, qua subject of experience, have a bodily existence, then the conservative distinction between the inner mental life and the mere colorless physical body is starting to erode, at least in my own case. And it seems legitimate to ask whether, if I cannot coherently view myself as a mind or a brain comfortably tucked away inside a skull, there is then any good reason to suppose that other minds would have to be thus tucked away from their moving bodies. Merleau-Ponty puts the point in this way:

> If I experience this inhering of my consciousness in its body and its world, the perception of other pefople and the plurality of consciousnesses no longer present any difficulty. If, for myself who am reflecting on perception, the perceiving subject appears provided with a primordial setting in relation to the world, drawing in its train that bodily thing in the absence of which there would be no other things for it, then why should other bodies I perceive not be similarly inhabited by consciousnesses? If my consciousness has a body, why should other bodies not 'have' consciousnesses? (Merleau-Ponty 1962, 351)

It is worth emphasizing that this is not a version of the argument from analogy. In fact, it is an indirect attack on the conservative assumption at the heart of the argument from analogy: the assumption that the mental lives of others must be concealed "behind" the body. For if I am essentially not thus concealed, why should I assume that others must be?

Note also that this type of argument lends no support to the behaviorist reductive view. For we started out by taking it for granted that I am a subject of perceptual experience, and thus have mental phenomena. I *see, touch*, and *hear* things; I don't just make noises and move about. The argument states that if we want to make it intelligible that I can see things, then we must grant that I also have to be able to move about. But "moving about" here becomes something an embodied subject, a minded creature does; it is not the "colorless movement" of a mere physical object.

The direct response involves a phenomenological thematization of others as we actually experience them. Thus, whereas the indirect response tried to show that since I cannot view myself as a mind hidden inside a body, there is little reason to expect others to be so, the direct response aims to show that indeed we do not experience other bodies as mere external shells that may at best license inferences as to the mental life "inside" them. Max Scheler points out that

> we certainly believe ourselves to be directly acquainted with another person's joy in his laughter, with his sorrow and pain in his tears, with his shame in his blushing, with his entreaty in his outstretched hands, with his love in his look of affection, with his rage in the gnashing of his teeth, with his threats in the clenching of his fist, and with the tenor of his thoughts in the sound of his words. If anyone tells me that this is not 'perception', for it cannot be so, in view of the fact that a perception is simply a 'complex of physical sensations', and that there is certainly no sensation of another person's mind nor any stimulus from such a source, I would beg him to turn aside from such questionable theories and address himself to the phenomenological facts. (Scheler 1954, 260)

Wittgenstein makes essentially the same point:

> Consciousness in another's face. Look into someone else's face, and see the consciousness in it, and a particular *shade* of consciousness. You see on it, in it, joy, indifference, interest, excitement, torpor, and so on. The light in other people's faces. Do you look into *yourself*

in order to recognize the fury in *his* face? It is there as clearly as in your own breast. (Wittgenstein 1967, §220)

Scheler and Wittgenstein argue that it is something like a philosophical (or perhaps pre-philosophical) *myth* that tempts us to place the mental phenomena of others outside our cognitive reach. If we look at our actual experience, a completely different picture emerges. We may *see* another's anger in her flushed cheeks, frowns, outbursts, and clenched fists. We don't need to link certain colorless physical movements to our own feelings, or indeed invoke some explanatory theoretical framework, in order to determine another's mental state.

That these points contradict the conservative accounts is obvious. But there are several reasons why they are equally antagonistic to the reductive accounts. First, in saying that we perceive these mental phenomena, the phenomenologists do not want to say that we perceive a person's anger the way we perceive "the color of the carpet" (e.g., Merleau-Ponty 1962, 184). The phenomenologists' interpretation of the observable, bodily or expressive phenomena is very different from the behaviorist interpretation. In fact, the supposedly merely external, physical occurrences that the behaviorist would single out as observational data may be hard for us to describe as such, without recourse to the mental phenomena. As Wittgenstein puts it,

> "We *see* emotion." – As opposed to what? – We do not see facial contortions and *make the inference* that he is feeling joy, grief, boredom. We describe a face immediately as sad, radiant, bored, even when we are unable to give any other description of the features. – Grief, one would like to say, is personified in the face (Wittgenstein 1980, §570).

Can one, except in unusual circumstances, give a very accurate description in purely physical or geometrical terms, of the features of an angry person's facial contortions? We may perhaps notice the color of her cheeks, the trembling hands, the frowns, but hardly much more than that; and often, indeed, we do not notice many of these features, as something else entirely dominates the scene: the person's foaming rage. To put it differently, another person's body is generally not perceived as a physical thing – as a *Körper*, to use a German expression invoked by some of the phenomenologists. Rather, it is perceived as a lived, expressive, or "animate" body – a *Leib* (e.g., Scheler 1954, 218).

The second point is a certain radicalization of the first. It isn't strictly correct to say that what we perceive are the "lived" or "expressive" bodies of others. This still retains too much of the idea common to the conservative and reductive accounts. As if the purpose was merely to slightly modify the conception shared by the latter, while accepting their point that we can never observe anything except "bodies". In contrast, the phenomenologists would insist that what we normally perceive are *bodily subjects*, unities that have a fundamental status and cannot be reconstructed by piecing together supposedly more basic elements. In Scheler's words: "Our immediate perceptions of our fellow-men do not relate to their bodies (unless we happen to be engaged in a medical examination), nor yet to their 'selves' or 'souls'. What we perceive are *integral wholes*" (1954, 261). The same point is expressed by Wittgenstein when he denies that it is a body or a mind that feels pain,

and claims instead that it is a "living human being" (and what resembles it) that has sensations, emotions, and the rest (Wittgenstein 1963, §§286, 573, 281). We perceive many varieties of inanimate physical objects: rocks, trees, and furniture, for example. But we also perceive animals and persons, and the phenomenologists insist that the latter are integrated wholes that it would be wrong to attempt to reconstruct in terms of something like a mind "inhabiting" an object of the former sort.

Finally, note one striking consequence of the behaviorist outlook. It postulates an epistemic symmetry between my relation to my own "mental life" and my relation to another's "mental life".[6] In both cases, my cognitive access goes via observable behavior. As Russell writes, the behaviorist

> maintains that our knowledge of ourselves is no different in kind from our knowledge of other people. We may see *more* because our own body is easier to observe than that of other people; but we do not see anything radically unlike what we see of others. (Russell 1949, 29; cf. Ryle 2000, 149)

Indeed, if the mental is nothing but behavior, I obviously have nothing else to go by when figuring out a person's mental state, whether that person is myself or someone else. But is this very convincing? Is it my expressive behavior that tells me that I am in pain? How well placed am I, generally, to notice that behavior – for example, my facial expression? To take it to extremes, do I look at my own eyes to determine what I am looking at? Obviously, this is absurd. True, I often have to pay attention to my own actions to figure out what my motives, desires, and even emotions (if they are complex enough) are; but to say, with Russell's behaviorist, that this is my only access to my mental life is simply absurd.

The phenomenological account of the mental therefore aims to preserve as fundamental the notion of an epistemic self-other *asymmetry*. To be in pain or feel sad is not a kind of perceiving that a person is in pain or feeling sad. I may perceive another's sadness – it may be visible in her facial expressions, for example – but I feel, and do not perceive, my own sadness. *Pace* the behaviorist, these are different *kinds* of access; but *pace* the conservative, they need not be interpreted as different *degrees* of access. Another person's sadness may just completely evident to me, but it is something I see (e.g.), not something I feel – although it may trigger emotional responses in me.

This final item in the phenomenologist's account is an extremely important one, because it helps us see something that motivates the conservative views, and it also makes it intelligible that the problem of other minds should continue to be construed as a skeptical problem. All the phenomenologists' talk about expressive bodies and integrated wholes notwithstanding, there is, after all, nothing that would count as having the same *kind* of access to another's mental life as she herself has. I cannot think her thoughts or feel her pains, and here, it may be felt, is the very root of the other minds problem. And so, in a sense, it is. But it is essential to realize that the

---

[6]The best explanation argument and at least some versions of theory–theory work with a similar symmetry: mental states as such (whether mine or someone else's) are unobservable theoretical entities.

kind of access to others that we are denied is one it doesn't make sense to suppose we might have. For if I did indeed feel "another's" sadness, think his thoughts, see the world through his eyes, and so on, then, as Husserl emphasizes, the other's mental life "would be merely a moment of my own essence, and ultimately he himself and I myself would be the same" (Husserl 1995, 109). In other words, the felt sadness wouldn't, after all, be another's sadness: it would be mine. Thus, no possible access to another mind is being envisaged here. So here we see the legitimate point underlying the skeptical and conservative approaches to the problem of other minds; but we also see that it is a point that need not worry us too much (the alternative turning out not to be an alternative after all). It is crucial, then, that we hold on to both of the following points: another's mental life is not essentially hidden and unobservable; but it is nevertheless in principle revealed to me in a way that differs from the way in which my own mental life is "revealed" to me.[7]

Concluding Remarks

I have suggested (1) that the traditional responses to the problem of other minds are inadequate because they fail to offer a radical enough revision of the Cartesian conception of mind and/or body; and (2) that phenomenology offers precisely such a revision. Unlike the conservative and reductive approaches, the phenomenological analyses of embodiment and expression, among other things, suggest a picture of the mental that makes intelligible social cognition.[8]

References

Avramides A (2001) Other minds. Routledge, London
Ayer AJ (1990) Language, truth and logic. Penguin Books, Harmondsworth
Carruthers P (1991) Introducing Persons. Routledge, London
Churchland PM (1988) Matter and consciousness, Revised edn. MIT Press, Cambridge, MA
Costall A (2006) Introspectionism and the mythical origins of scientific psychology. Conscious Cogn 15:634–654
Dennett DC (1971) Intentional systems. J Philos 68:87–106
Descartes R (1984) Meditations on first philosophy (trans: Cottingham J). In: The philosophical writings of Descartes, vol II, Cambridge University Press, Cambridge, pp 3–62

---

[7] For discussion of these points, see Overgaard 2007, chapters 5 and 7.

Gallagher S, Zahavi D (2007) The phenomenological mind. Routledge, London

Gallese V, Goldman AI (1998) Mirror neurons and the simulation theory of mind-reading. Trends Cogn Sci 2:493–501

Goldman AI (2000) Folk psychology and mental concepts. Protosociology 14:4–25

Gopnik A (1993) How we know our minds: the illusion of first-person knowledge of intentionality. Behav Brain Sci 16:1–14

Gopnik A (1996) Theories and modules; creation myths, developmental realities, and neurath's boat. In: Carruthers P, Smith PK (eds) Theories of theories of mind. Cambridge University Press, Cambridge, pp 169–183

Gordon RM (1996) Radical simulationism. In: Carruthers P, Smith PK (eds) Theories of theories of mind. Cambridge University Press, Cambridge, pp 11–21

Hobson RP (1991) Against the theory of 'theory of mind'. Br J Dev Psychol 9:33–51

Hull CL (1943) Principles of behavior: an introduction to behavior theory. Appleton-Century-Crofts, New York

Husserl E (1989) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: second book (trans: Rojcewicz R, Schuwer A). Kluwer, Dordrecht

Husserl E (1995) Cartesian meditations: an introduction to phenomenology (trans: Cairns D). Kluwer, Dordrecht

Hyslop A (1995) Other minds. Kluwer, Dordrecht

Leslie AM (1987) Children's understanding of the mental world. In: Gregory RL (ed) The Oxford companion to the mind. Oxford University Press, Oxford, pp 139–142

Leudar I, Costall A (2004) On the persistence of the 'problem of other minds' in psychology. Theory Psychol 14:601–621

Merleau-Ponty M (1962) Phenomenology of perception (trans: Smith C). Routledge, London

Mill JS (1979) In: Robson JM (ed) An examination of Sir William Hamilton's philosophy: collected works, vol IX. University of Toronto Press, Toronto

Noë A (2004) Action in perception. MIT Press, Cambridge, MA

Overgaard S (2007) Wittgenstein and other minds. Routledge, New York

Overgaard S, Zahavi D (2009) Understanding (other) minds: Wittgenstein's phenomenological contribution. In: Levy D, Zamuner E (eds) Wittgenstein's enduring arguments. Routledge, London, pp 60–80

Russell B (1949) The analysis of mind. Allen & Unwin, London

Ryle G (2000) The concept of mind. Penguin Books, Harmondsworth

Sartre J-P (1989) Being and nothingness: an essay on phenomenological ontology (trans: Barnes HE). Routledge, London

Scheler M (1954) The nature of sympathy (trans: Heath P). Routledge and Kegan Paul, London

Searle J (1983) Intentionality: an essay in the philosophy of mind. Cambridge University Press, Cambridge

Wittgenstein L (1963) Philosophical investigations (trans: Anscombe GEM). Blackwell, Oxford

Wittgenstein L (1967) In: Anscombe GEM, von Wright GH (eds) Zettel (trans: Anscombe GEM). University of California Press, Berkeley

Wittgenstein L (1980) In: von Wright GH, Nyman H (eds) Remarks on the philosophy of psychology, vol II (trans: Luckhardt CG, Aue MAE). Blackwell, Oxford

Mutual Gaze and Intersubjectivity[1]

Beata Stawarska

In this paper I will address the role of mutual gaze in social cognition. The exposition will run in two steps. First, I propose to examine some recent studies of joint visual attention in order to substantiate the view that social cognition is operative in infancy prior to the emergence of theoretical skills required to make judgements about other people's states of mind. Such social cognition does not depend on intellectual procedures but rather on the communicative potential inherent in human bodies, for example, the ability to directly engage with others via mutual gaze. This view of sociality as dependent on bodily practices is broadly consistent with the phenomenological philosophy of Merleau-Ponty, which I propose to address second. Specifically, I will examine vision in the context of reversible dynamics which Merleau-Ponty believes regulate intercorporeal relations. This will allow me to expose some inner difficulties within Merleau-Ponty's position as well as to point out the ways of resolving them by means of combined insights from developmental psychology and the analyses of interpersonal connectedness drawn from the dialogic tradition in philosophy.

Mindsight

Research of joint visual attention has recently become entrenched in the theoretical framework of the Theory of Mind (ToM) approach to social cognition, following the influential mindreading model developed by Simon Baron-Cohen (2001). Broadly speaking, this approach stipulates that we need to read mental states into the overt behavior of others in order to engage with them as intentional subjects, analogous to ourselves. The perceptual skill of attending to another person's gaze and recognizing whether or not she is looking at me at a given moment plays an

B. Stawarska
Department of Philosophy, University of Oregon, Eugene, 97403-1295, OR, USA

[1] This chapter is a revised version of my earlier piece "Mutual Gaze and Social Cognition." *Phenomenology and the Cognitive Sciences*, vol 5.1, pp. 17–30, 2006.

important role in this theoretical framework, insofar as the ability to detect the eye direction of others provides one of the four building blocks of a fully fledged ToM apparatus. The eye direction detector (EDD) is argued to be an innate module responsible for detecting eye-like stimuli, identifying where a person looks and interpreting gaze as an act of seeing, which the infant is said to do by analogy with its own experience. EDD allows the infant to establish dyadic relations with its caretaker through shared gaze. The correlated abilities of detecting eyes, following their line of regard and interpreting their activity as *seeing* do not, however, suffice to enable contact with another person's mind on this model. Should our social cognition be limited to these visual techniques, we would inhabit an autistic-like universe, i.e., be devoid of effectively knowing others as mindful (see Baron-Cohen 2001, 44). Detecting eye direction becomes significant for mindreading only at a developmentally later stage (roughly 9 months) when the child begins to share other people's visual representations of the world by following their line of regard. These shared attention mechanisms (SAM) which triangulate between self-other-world, are said to trigger the ability to theorize about other people's unobservable states of mind or to build a ToM.

As the reader may know, the ultimate interest of Baron-Cohen's research is to address the disturbances of social relations in the autistic population and to respond to the question whether autistic subjects can read the mental states of others and therefore have a ToM. The so-called false belief task was decisive in settling this issue. The task typically involves a subject whose knowledge of a given state of affairs, e.g., the contents of a Smarties box, surpasses that of a third party who is temporarily absent from the experimental situation and whose beliefs regarding the given state of affairs the subject is asked to predict.[2] It was found that autistic subjects fail to predict false beliefs; having discovered upon inspection that the Smarties box contains a bunch of pencils, autists predict that the third party will believe likewise rather than attributing the appropriate albeit false belief to them (that the box contains Smarties) – the belief that they themselves held in the initial stages of the experiment. Insofar as the autist's prediction fails to allow for beliefs other than the subject's own present beliefs, the false belief task has been regarded as proof that an autist cannot access other minds or that her experience is characterized by a pervasive *mindblindness*.

It may be surprising that the metaphor applied for impeded social cognition is borrowed from the domain of vision; after all, visual perception does not permit us to connect minds on the ToM model. Alternatively, we may explore the hypothesis that normally functioning social life does depend on the perceptual ability to *see* other minds or, to coin an appropriate term, on *mindsight* (as far as the sighted but not blind population is concerned). Clearly, this line of thought would deviate from the ToM paradigm of social cognition; however, some recent studies validate this alternative view of visual social cognition as mindsight or direct perception of mental states of other people already within dyadic relations. Support for this view

---

[2]False belief tasks were originally designed by Wimmer and Perner (1983).

can be found in the recent research by Juan Carlos Gomez, which I propose to examine.

Gomez combines experimental work in primatology with insights drawn from developmental psychology to outline the communicative strategies of joint visual attention, recruited by non-human primates as well as preverbal infants, to actively engage with their social milieu. Since my present focus lies in human sociality, I will limit the discussion about joint visual attention to the field of developmental psychology. For the infant, joint visual attention with the adult initially takes the form of direct eye contact or mutual gaze before giving rise to a more complex formation which alternates between attending to each other's eyes and jointly intending an object present in their visual field (a toy, food). Sustained focus on the caretaker's eyes has been found in infants as young as 1 month of age (Wolff 1963), and it has been documented as a basic trait of a normally developing mother–infant relation. The focus on the eyes is made possible by the relative functional maturity of the infant's visual-motor system, attained already by the third month, while other motor behaviors are less mature. This maturity permits the infant to visually engage with others in face-to-face situations *almost* as an equal partner; or as Stern puts it, "the gazing patterns between a mother and her infant of this age constitute the first dyadic system in which both members have almost equal control over and facility with the same behavior" (Stern 1974).

The emergence of eye contact regularly triggers positive social responses of infantile smiling. It also dramatically affects the mother's behavior, giving her the sense of connecting to the baby as a person for the first time (Wolff 1963; Robson 1967). One may wonder why there should occur such a dramatic change in the infant-caretaker relation once their eyes have met. To explain this it is important to note, as Gomez (1994) does, that looking at the eyes of a person who in turn is looking at one's own eyes is different from looking at any part of that person's body as it moves and gestures. Sheer observation of another human body does not yield the kind of self-other relation that is instantiated by eye contact. For in eye contact you not only observe the eyes of the other person but are also checking her attention, while the other person who attends to your eyes is checking your attention as well. In other words, the other is attending to your attention while you are attending to hers or "she is doing in relation to us what we are doing in relation to her." Gomez (1994, 72/73) proposes to call this mutual attention to each other's signs of attention in shared gaze "attention contact." It is a process of social attunement.

Attention contact relies on addressing the other's attention directly or on a *second-person relation* with the other, whereas observation of the body without contact consists, on the other hand, in a *third-person orientation* to the other as a thing. This distinction between second and third person relations can be rephrased in terms of the I/Thou and the I/it relation, such as it figures in the philosophy of dialogue developed by Martin Buber (1958). In the I/Thou relation, the other is directly experienced as a living and conscious fellow being. This experience is nonpredicative, i.e., not based on conscious judgement cast in an "X is a minded being" form, but rather on unmediated awareness of being in the presence of another person. I do not infer or theorize that the other person is likely "minded."

Nor do I need to have access to the contents of her mind to regard her as a person or to be aware of her awareness. In the I/Thou relation I stand to the other in a relation of mutuality, of addressing the other and responding to the other. It is this mutual quality of the I/Thou relation which forms the axis of direct social relationship and distinguishes it from a third-person mode of relation to the other which is unidirectional and lacks mutuality.

We realize then that the relation to an embodied other may follow two non-identical orientations which are determined by *where we look*: looking at the eyes, when reciprocated by the other, enables an I/Thou relation where the self and other regard each other as subjects, while looking at the body produces an I/it relation where the other is regarded as a non-participatory element of the visual field. We then begin to understand the dramatic change brought by mutual gaze to the mother–infant relation: as a relation involving mutuality, eye contact transforms the mother's third-person orientation to the baby *at which* she looks and *of which* she needs to take care to a second-person relation *to* a living, attentive and responsive human being *with whom* direct social ties can be drawn. We realize also that embodied social relations need not have an objectifying or third-person orientation to the other as a physical body and that social relations are not preconditioned by an already formed body image, as the mirror stage theory claims.[3] If mutuality provides the key to direct social relations, then the objectifying approaches devoid of mutual awareness do not provide a paradigm for intersubjectivity.

Returning to my earlier discussion of how Gomez' work provides an alternative view to the theory of mind approach to social cognition, note first of all that attention contact obtained via mutual gaze is not the result of complex metarepresentational abilities. To be sure, the reflexive structure it involves is comparable to the entangled effect produced by two mirrors facing each other (A reflecting B reflecting A reflecting B, and so on – or A attending to B attending to A attending to B, and so on). As such, attention contact invites a comparison to a cognitive mechanism of imputing second and higher order thoughts to the other person. Human adults, especially philosophers, are able and likely to translate this attentional loop occurring in eye contact into a chain of thoughts accessed via our inferential powers. However, infants are devoid of these metarepresentational powers required by the theory of mind. Even though they may have direct or first order representations of their milieu, they are unable to have representations of other people's representations of objects, persons or events. Hence, the reflexive structure of mutual visual contact is not a product of analytic exercises but rather is engendered by the lived experience itself.

Gomez theorizes eye contact as a perceptual rather than an intellectual process. Visual attention is not an invisible mental process to be inferred but is directly perceived in the other person's overt behavior via external signs of attention such as the direction of the gaze as well as facial expressions. In the sighted population, eye contact serves to read or to *see* the minds of others in direct mutual interaction.

---

[3]See Stawarska 2004 for further discussion of the mirror stage and the difficulties inherent in the theory of social cognition based upon it.

It is valid therefore to uphold the category of *mindsight* as direct visual access to other's mental states obtained in the mutual attention loop of reciprocal gaze. As a perceptual rather than purely cognitive process, mindsight is firmly rooted in embodied human capacities such as gaze tracking and appreciating the other's facial expression. Furthermore, it can be concluded that, as far as the sighted population is concerned, the experience of mutual gaze serves the formative function of establishing social bonds by directing the infant's and adult's attentional states onto each other and locking them in mutual awareness. For even though the gaze is not the exclusive way of attending to other people (the blind population does not use *visual* forms of attention) and even though mutual gaze is not the only form of attention contact, still the gaze appears as the primary and privileged way to orient attention and to connect with other persons afforded by sighted bodily organisms. Unlike touch, it does not depend on immediate proximity to transmit the signs of attention and elicit contact; gaze enables, therefore, a broader range of opportunities for mutual attention than touch does. The former also precedes hearing in constituting permanent identities of persons – infants learn to locate a person via the voice later than they do visually. The gaze has, therefore, a developmental priority over the auditory sense in connecting people. The broad range of contact possibilities and developmental primacy bestow a privileged position onto the visual sense in the field of social relations. To rephrase this point in more general terms, bodily sensibility and the affordances of human bodies, such as the affordance of sighted bodies to engage in direct contact via mutual gaze, actively shape the kind of relations that obtain between minds. Not only may it be possible to see the other's mind in her body, but the body and its sensibility actively inform the specific ways in which the minds interrelate.

The conception of mindsight proposed here could be used to investigate social difficulties within the autistic population from a different, and possibly revealing, angle. By investigating the *broadly construed* perceptual abilities of the autists, which include not only negotiating the physical environment via the visual sense but also registering and responding to the manifestations of other persons' feelings, whether through facial expressions, gaze, body posture, gait or gesture, it could be found that the challenge faced by an autistic individual lies at a cognitively 'lower' level than has been assumed in the theory of mind model. The challenge may lie at the level of *seeing* which extends beyond the narrow physicalistic definition of vision in terms of an intake of meaning-neutral sense data to encompass the semi-instantaneous as well as nuanced ability to perceive the other person's state of mind in her face and body, prior to or independently of launching complex theoretical mechanisms for explaining and predicting behavior.

Some experts in autism who transcend the narrow cognitive mechanism based approaches to the problem and endorse the formative role played by the social milieu in infancy and early childhood in the development of autistic traits have considered such an alternative conception of *seeing minds* in their research. Peter Hobson (2004) clearly recognizes the importance of a well functioning visual sense for forging social ties and has extensively studied the disastrous effects of the depravation of sight-dependent interactions with others, such as social smiling, eye contact,

and eye tracking in cases of congenital blindness for the social development of infants and children. Blind infants tend to be withdrawn and indifferent to the social environment to an extent that prompted Hobson (1997) to classify their condition as *autistic-like*. Needless to say, the two populations suffer due to different kinds of depravation. In the case of blindness, "the inability to see other people's relations with themselves and the world plays a critical role" (Hobson 2004, 192). In the case of autism, it is the inability to perceive other people as having feelings, to respond to them with feelings of one's own, as well as to identify with attitudes of others directed at events and objects in the world that is critical. Hobson notes that these latter difficulties have to do with 'seeing', "but in a sense that extends beyond vision" (Ibid.). It could be added, however, that human seeing typically extends beyond vision to include an affective component, and that we perceive people as well as objects and events *with feeling*. If that is the case, the comparison between the blind and the autistic populations can be rephrased in the following terms. In the former, the perceptual foundation upon which to build the structure of social relations is missing; in the latter, the perceptual foundation is present but impoverished or affectively 'neutralized' in a way that leads to a socially anaesthetized view of the world. The task open to autism research is to investigate this pathology at the foundation level of perception and affectivity, rather than at a cognitively higher level of theory, as tends to be the case in current debates on social cognition. The key question to ask is why seeing has become dissociated from feeling or affectively impoverished in the autistic group in a way that constrains the typical ability of mindsight.[4]

Double Sight

When discussing the embodied foundation of social relations, one typically turns to the phenomenological philosopher Maurice Merleau-Ponty for illumination. Merleau-Ponty consistently developed the idea that the dynamics of embodiment provides the clue to our immediate experience of others, whom we discover with a similar degree of intimacy and familiarity that typifies the experience of our own body (see Merleau-Ponty 2000, 135, 138). The experience of the other depends on and employs the corporeal dynamics which are already at work within the borders of the body proper and which can be generalized onto interactions with other bodies. Or, seen from another perspective, my own body and the body of the other are both instances of the same corporeal process that runs throughout the sensible/sentient world, and which can be subsumed under the heading *reversibility*. A typical example

---

[4] The category of mindsight I have referred to may be substantiated philosophically by Wittgenstein's comments regarding the human ability to directly *see* (rather than infer) other minds (see Overgaard 2006) for further discussion. The author makes an observation similar to my own that autism may consist in a perceptual rather than theoretical inability (so-called aspect blindness, i. e., inability to perceive emotional significance of things and people).

of this reversible process is found in the case of my two joined hands alternating between the active modes of touching and passive mode of being touched. Here my body becomes revealed to me as living flesh responding to touch by touching back in turn. The reversal between actively touching and being touched can continue as long as a sensitive organism like myself may wish to engage in this reflexive play, and yet at no time will the experience of actively touching coincide with the feeling of being touched in the same organ. The active and passive modes are conjoined in a crosswise or *chiasmatic* fashion, which prevents them from overlapping. This chiasmatic structure of a reversible process explains why it can, following Merleau-Ponty, regulate interpersonal relations as much as it regulates the intrapersonal phenomenon of double touch. An analogous case of a reversal between touching and being touched occurs between two bodies, as in the case of a handshake where I touch the other who touches me as well. When shaking hands with another person, his hand becomes 'annexed' into the reversible system that I previously discovered operative within my own sensible body (Merleau-Ponty 1964a, 168). Hence the general *modus operandi* of reversibility apparently remains the same when it gets instantiated on the *inter-* rather than the *intra-*corporeal level.[5]

Consider the similar structure of the handshake example and mutual gaze discussed earlier. Both rely on mutuality to establish direct contact between the self and the other within a given sense modality, by reciprocated touch or sight. If touching the other and being touched by her can be interpreted along the lines of attending to the other and being the focus of the other's attention, then we can construe mutual touch as a case of attention contact similar to, even though not identical with, eye contact. This similarity would explain why mutual touch and mutual gaze produce an analogous effect of directly connecting with the other as a person, whether or not the bodies meet. It would explain also why mutual gaze, even though it disperses across the minimal distance necessary to see, may share the focus, intensity and affective charge brought by tactile proximity with another person. We are then brought back to the importance of mutuality for establishing direct social interaction, previously subsumed under the I/Thou relation, and realize that our bodies afford mutual attention within the sense modalities of both sight and touch. It seems that these two sense modalities afford direct mutual relations because they conjoin the two communicative functions of initiating and responding to contact within one and the same sensible organ. The eyes which see and can be seen, the hands that touch and can be touched are equipped and predestined for direct social interaction thanks to the double sender/receiver role they play in situ. (It appears that the blind population relies primarily on touch to obtain such immediate interaction or "attention contact.")[6]

The question is whether the visual sense can be interpreted along the lines of reversibility, specifically whether there is a direct analogy between intracorporeal and intercorporeal experiences in vision that Merleau-Ponty argued obtains in the

---

[5] See Merleau-Ponty (1968, especially p. 142), (1964a, p. 168), (2003, p. 109). See also Barbaras (2001, esp. pp. 277–305). For a critical discussion of this view, see Stawarska 2002.

[6] See John M. Hull's (2001, pp. 51/2) penetrating comments about how certain organs are prewired for communication.

modality of touch. The intracorporeal experience of seeing my body plays a formative role in the process of discovering sensible reversibility in the modality of vision for Merleau-Ponty, comparable to the aforementioned experience of double touch.[7] For if I had a body constructed in such a way that I could not touch *or see* myself, then this body of mine would be "not really flesh, not really the body of a human being" (see Merleau-Ponty 1964b, 163). To qualify as flesh, the human body needs to reflexively turn to itself via its reversible senses. One may wonder, however, whether looking at oneself really serves as the candidate for the discovery of the reversibility of the active and passive modes of sight. If it is accepted that reversibility stands for the double experience of seeing and being seen, then it can hardly be claimed that observing the façade, the sides of and the lower backside of my body could ever be a source of this double experience. It seems that my solitary body, even though it is visible and I do see a substantial part of it, does not produce the lived experience of being seen as a complement to me seeing it. Unlike the organ of touch which consists of the entire epidermal surface of the body and which affords a multitude of reflexive intrabodily experiences of rubbing, scratching, holding, squeezing, pinching, caressing, etc., that may be enacted not only by the limbs but also by the mouth, the organ of sight is fixed in the two openings within the upper half of my face which act in synergy to produce binocular vision but do not engage in reciprocal intrabodily relations. The body proper does not generate reciprocated looking (even though it is the source of double touch). Even though I see and my body is seen, still my activity of seeing is not complemented by the passive mode of being seen when I look at myself.

Now, what happens when we extend the case of looking at the body to two actors and engage them in reciprocal observation of their bodies? Would this count as a case of reversibility within the modality of vision, as a co-presence of seeing and being seen? It seems not, insofar as reciprocated seeing occurs only when two actors look each other in the eyes and not at their bodies. Recall the previously introduced distinction between the I/Thou and the I/it relation: the co-presence of seeing and being seen occurs only within the mutual visual relation between I and Thou and does not obtain within the relation between I and it which lacks mutuality. In the latter case the other looks at my body and I look at hers, yet despite this temporal co-presence of other-oriented acts, reciprocated looking is precluded by the it-orientation adopted by both actors. To cast the argument in Alfred Schutz's (1967) terms, we need to distinguish between social *observation*, where two partners observe each other, from social *relationship*, where two partners mutually engage each other.[8] Even though both types of relation typically occur in the modality of vision, it is the latter type only that exhibits reciprocity and realizes the dual relation of seeing and being seen that Merleau-Ponty attributed to reversibility.

Following Schutz's lead further still, reciprocity is not an additive unity composed of two sighted persons' simultaneous observational acts directed at each

---

[7] See, e.g. Merleau-Ponty (1968, 146) on the body seeing itself as an example of reversibility.
[8] For this distinction, see chapter 4 of Schutz (1967).

other. We cannot therefore derive social relations from social observation. Reciprocity is a single non-composite (and non-objectifying) act realized in equal measure by the two actors. In Schutz's words, it is "one undivided intentional Act" which typifies face-to-face relations (1967, 168).

Let the following anecdote illustrate that point. A child covers over his eyes and once they are firmly closed solemnly declares to another child: "you cannot see me." The other child looks on, fascinated. An adult witnessing the scene would probably find it highly amusing. I believe, however, that it provides a good source of insight into the unified or non-composite nature of mutual visual attention, as distinguished from the observation of the body in a third person mode without reciprocity. Only the mutual attention experience is annulled as soon as one of the parties stops looking: I have my eyes closed, hence I am not seeing you and you are not seeing me looking at you – hence "you cannot see me." Needless to say, the figure of my body remains in clear view whether or not the eyes are closed and that is why an adult witness may find the child's declaration amusing. Yet it remains true that the pact of shared visual attention gets broken as soon as one of the parties covers over his eyes – and so you cannot see me seeing you seeing me, and so on. The child pronounces only the first clause from the long chain that composes the attentional loop. Still, he refers to the phenomenon of mutual awareness that gets enacted via shared gaze and which, I believe, provides a complete experience of conjoined seeing and being seen or, to coin an appropriate term, of *double sight*.[9]

If that is the case, we may conclude that the intercorporeal experience of *reciprocity* provides a condition of possibility of *reversibility* within the modality of vision. Sight is a socially informed sense modality which becomes fully fledged, i.e., reversible, via interaction with other seers in mutual visual attention. The other person plays a constructive role in the process – she co-enacts the reversibility of vision with me. We are therefore led to argue for the transformative potential of interpersonal experience in the domain of vision, not necessarily or exclusively along the lines of the objectifying effect attributed to the other's gaze by Sartre (1956), but rather along the lines of the affiliative gaze that that may bind two individuals in eye contact. Referring back to the discussion of autism, I hypothesize that an autistic individual does not attain the reversible dimension of sight due to lack of requisite visual social experience, such as mutual gaze. The autist may be able to navigate the visual space correctly but it is not a space that is affectively shared with others.

The above analysis serves to place Merleau-Ponty's account of reversibility in a critical perspective. It turns out to be impossible to derive the reciprocal visual relation between self and other from a dynamic operative within their solitary bodies. The lived experience of double sight could not have been realized independently by either of the actors it co-involves. Insofar as, as I argued earlier, looking at oneself does not add up to a lived experience of conjoined seeing and being seen, then there is no reversibility operative within the modality of vision on the intracorporeal

---

[9] Compare again Hull (2001, 51).

plane, and it is the intercorporeal experience of mutual gaze that installs reversibility for the first time rather than replicating a preexistent process. Rephrased in terms taken over from Buber, the novelty of the intercorporeal experience of mutual gaze consists in its establishing an I/Thou type of relation within one's embodied experience for the first time, insofar as a solitary body seeing itself seems capable of a unidirectional I/it type of relation only. It follows that we should not use a uniform category of reversibility to cover both types of relations but rather preserve a distinction between them within any theory that strives to do justice to the specificity of social life.

This objective can be attained by integrating contributions from the dialogic tradition in philosophy into the classical phenomenological accounts of embodiment. Consider Buber's notion of I–you connectedness in this context. Buber's philosophy strategically opposes the living dialogic relations of first-to-second person type to the detached spectatorial stance deployed by an external observer. Buber typically casts this opposition in verbal terms, as a dichotomy between two kinds of basic words (*Grundword*): *I–you* and *I–it* (1958, 53). He is therefore in general agreement with the notion embraced by linguists like Benveniste (1971), that the first and second person pronouns need to be distinguished from the third person one. Benveniste contended in fact that only the former two can justifiably be termed *personal*, while the third designates a non-personal entity, which could refer to a thing or a corpse. In this he followed the definitions used by Arab grammarians, for whom the first person designates 'the one who speaks', the second 'the one who is addressed', while the third 'the one who is absent' (197). The third person pronoun lacks what the other two pronouns capture, namely positive participant involvement in discourse. That is why the referents of the third-person pronouns may be seen as 'non-persons', for their position in the speech act is defined in exclusively negative terms, in contrast to the referents of *I* and *you* (Lyons 1977, 638). Contrary to *I* and *you*, the third person pronoun referents are not defined in terms of speech roles. Furthermore, in accordance with the Arab classification, while the referent of the third person pronoun may be absent and yet identifiable, in order to identify the persons indicated by *I* and *you*, one needs to be present in flesh and blood in the event of discourse. *I* and *you* designate therefore the interrelated speaker and listener roles occupied by the interlocutors in the context of utterance; they mark the self and the other in the relation of mutual address, in speaking and listening *to* one another. *I* and *you* are therefore intra-discursive, whereas the relation between an *I* and an *it* extends to an extra-discursive entity, a non-participatory third party spoken *about* and which is not called upon to respond, not even to listen.

Even though Buber adopts this discourse-based distinction between personal and impersonal pronouns, he does not limit the underlying attitudes or stances (*Haltung*) to discursive acts alone. The I–you relation, which he considers more basic and primary than the I–it, is primarily an embodied (*Leiblichkeit*) relatedness to the other embedded in the shared natural world. That relation can be thematized as speech, i.e., as the living utterance issuing from a mouth and terminating in an ear, which travels and resonates in the elemental air. Yet Buber regards a shared glance with another as a rightful instance of interpersonal connectedness in the

first-to-second person mode. Furthermore, the primacy of I–you within human sociality is construed as a strong developmental or genetic claim about the dual state of the embodied human who lives initially in an entangled organic interrelation with the m/other in prenatal life. The life in utero is for Buber an instance of the natural bond (*Verbundenheit*) with the mother, where there is a "flowing toward each other, a bodily reciprocity." (1970, 76). This intimate bond developed in utero remains pervasive throughout post partum life, not in the naïve sense of a craving to crawl back into the womb, but rather as an undying longing for one's true You. The interdependency of the infant and the mother captures therefore what Buber terms the a priori of relation, the yearning for an innate You, which predates the separation of birth (1970, 78). Buber regards this corporeal relational a priori as a precondition of speech. He terms it a "wordless anticipation of saying You" (1970, 78), an original orientation to the other which precedes and conditions the verbalized second-person address in the subsequent stages of human life.

I conclude therefore that Buber's notion of first-to-second person relations both precedes and exceeds the narrow linguistic construal of I–you as verbalized address, and that it encompasses a host of embodied reciprocal relations with others. It helps therefore to complement the classical phenomenological accounts of embodiment, and to make a strong case for a social rather than primarily individual conception of the lived body. Buber's thesis that the I–you dynamic is primary, both in the developmental sense, as well as in the sense of organizing direct interpersonal relations throughout human life, provides an important corrective to the discussions of sociality which privilege the I–it mode, and cover over the non-objectivating, participatory character of social relatedness in face-to-face encounters, which is more fundamental than any discourse *about* a third. To be sure, Buber's dialogic philosophy does not dispute the importance of indicative statements in the third person, nor does it suppose that I–you relationality could exist without an attachment or even an entanglement with the I–it orientation (1958, 69). However, in Buber's philosophy, one can only make statements about a non-participatory third on the base of preexisting interpersonal connectedness in the I–you mode. The opposite view would yield an unintelligible rendering of language as a sterile code without communicative involvement, akin to a perplexing vision of the social world as a series of disengaged minds thematized exclusively in terms of an unrelational it.

The perspective of distinct and co-dependent I–you and I–it relations borrowed from Buber can be concretized by the findings from developmental psychology, as discussed in the work of Gomez. For example, it could be shown that the infant's capacity of joint visual attention which alternates between mutual gaze and intending a visually present object is a complex phenomenon which alternates between direct dyadic mutuality and joint visual experience. This substantiates the point that the I/Thou relation established in eye contact is not limited to interlocking two sighted persons in a dyadic exchange of glances, at the exclusion of the surrounding world, but is rather a precondition of sharing and communicating about the world. After all, if my communicative gestures, such as pointing to something, are to reach their addressee, I need to make sure that she is paying attention to my intention to

communicate something in the first place. This intention to communicate is typi-cally conveyed via mutual gaze, which provides a firm base for transmission of information about objects, persons and events, within infancy as well as adulthood.

Combining insights from developmental psychology and the dialogic tradition in philosophy also help to refine the somewhat monolithic notion of the it-orientation in Buber's philosophy. The latter is not, after all, a uniform category but includes different variants of non-reciprocated social acts. Apart from regarding the other as a physical thing, it may include an agent orientation where the other is regarded in terms of her capacity to accomplish a given task in the world but where no mutual contact is established for the purpose. Such agent orientation can be exemplified by the behavior of autistic individuals who engage the other by taking her by the hand and leading her towards the site where a requested object can be found without addressing the other in a face-to-face mode. In this case, the other is engaged with no attention contact being established.[10] This is not to imply, however, that all types of social acts which exhibit the it-orientation are necessarily objectifying or autis-tic-like. For example, we do not necessarily objectify others in social observation and watching children at play is not (typically) an objectifying act![11] However, I would add that a failure to engage in other than observational and instrumental acts with others does denote a failure to fully recognize the other as a person. In Kantian terms, the other regarded as a means or mundane thing *only* is not grasped as a personal other.

Concluding Remarks

In this paper I examined the challenges that research into mutual gaze may raise to accounts of social cognition both within the ToM perspective and within Merleau-Ponty's philosophy. The former challenge, subsumed under the heading of mind-sight, is to the effect that it is possible to visually engage with other minds through eye contact and that eyes provide "windows" to our minds. The latter challenge, subsumed under the heading of double sight, is that such visual engagement plays a transformative role in embodied experience by installing reversibility in the modality of vision for the first time. We are brought to the realization that the sight modality both enables connecting with minds of others and is affected by others in the process.

The focus on mutual gaze enables me to point to the specific type of a direct second person relation where the other figures as person who is addressed in a face

---

[10]For a discussion of the other being engaged as an agent but not as a person or a subject by autistic individuals and gorillas, see Gomez (1991) and Gomez et al. (1993).

[11]As I am reminded by S. Overgaard.

to face situation rather than as an alter ego or another mind about whom I theorize in a third person mode. As such, it serves as a plea to include the second person perspective of dialogue within phenomenological as well psychological accounts of social cognition rather than rely primarily on the third person perspective where the other tends to be construed as a 'problem' to be resolved. The notion of the second person perspective, borrowed from the dialogic tradition in philosophy, captures the lived dynamic of mutuality in direct social relatedness and must be accommodated within any experientially based theory of social life.[12]

References

Argyle M, Cook M (1976) Gaze and mutual gaze. Cambridge University Press, Cambridge

Barbaras R (2001) De l'être du phénomène. Sur l'ontologie de Merleau-Ponty. Millon, Grenoble

Baron-Cohen S (2001) Mindblindness. An essay on autism and theory of mind. MIT Press, Cambridge

Benveniste E (1971) Problems in general linguistics. University of Miami Press, Coral Gables

Buber M (1958) I and thou. Scribners, New York

Gomez JC (1991) Visual behavior as a window for reading the mind of others in primates. In: Whiten A (ed) Natural theories of mind. Blackwell, Cambridge, MA

Gomez JC, Sarriá E, Tamarit J (1993) The comparative study of early communication and theories of mind: ontogeny, phylogeny, and pathology. In: Baron Cohen S, Tager-Flusberg H, Cohen D (eds) Understanding other minds: perspectives from autism. Oxford University Press, Oxford

Gomez JC (1994) Mutual awareness: a Gricean approach. In: Parker ST, Mietchell RW, Boccia ML (eds) Self-awareness in animals and humans. Cambridge University Press, Cambridge

Hobson P (2004) The cradle of thought. Oxford University Press, Oxford

Hobson P (1997) Are there "autistic-like" features in congenitally blind children? J Child Psychol 38(6):693–703

Hull JM (2001) On sight and insight. Oneworld Publications, Oxford

Jopling D (1993) The philosophy of dialogue. In: Neisser U (ed) The perceived self. Ecological and interpersonal sources of self-knowledge. Cambridge University Press, Cambridge

Lyons J (1977) Semantics, vol I & II. Cambridge University Press, Cambridge

Merleau-Ponty M (2003) La Nature. Northwestern University Press, Evanston

Merleau-Ponty M (2000) The prose of the world. Northwestern University Press, Evanston

Merleau-Ponty M (1968) The visible and the invisible. Northwestern University Press, Evanston

Merleau-Ponty M (1964a) The philosopher and his shadow. Signs. Northwestern University Press, Evanston

Merleau-Ponty M (1964b) Eye and Mind. The primacy of perception and other essays. Northwestern University Press, Evanston

Overgaard S (2006) The problem of other minds: Wittgenstein's phenomenological perspective. Phenomenol Cogn Sci 5(1):53–73

Robson KS (1967) The role of eye-to-eye contact in maternal-infant attachment. J Child Psychol Psychiatry 8(1):13–25

Rochat P (2001) The infant's world. Harvard University Press, Cambridge, MA/London

Sartre JP (1956) Being and nothingness. An essay on phenomenological ontology. Philosophical Library, New York

---

[12] See, e.g. Rochat (2001) on the necessary interrelation between mutuality and sociality.

Schutz A (1967) The phenomenology of the social world. Northwestern University Press, Evanston

Stawarska B (2004) The body, the mirror and the other in Merleau-Ponty and Sartre. In: Gallagher S, Watson S (eds) Ipseity and alterity: interdisciplinary approaches to intersubjectivity. Presses Universitaires de Rouen, Rouen

Stawarska B (2002) Reversibility and intersubjectivity in Merleau-Ponty's ontology. J Br Soc Phenomenol 33(2):155–166

Stern DN (1974) Mother and infant at play: the dyadic interaction involving facial, vocal and gaze behaviors. In: Lewis M, Rosenblum LA (eds) The effect of the infant on its caregiver. Wiley, New York

Varela FJ, Thompson E, Rosch E (2000) The embodied mind. Cognitive science and human experience. MIT Press, Cambridge

Wimmer H, Perner J (1983) Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13:103–128

Wolff PH (1963) Observations of the early development in smiling. In: Foss BM (ed) Determinants of infant behavior, vol II. Wiley, New York, pp 113–138

Knowing Other People's Mental States as if They Were One's Own

Frédérique de Vignemont

> *Take a sentence of a dozen words, and take twelve men and tell to each one word. Then stand the men in a row or jam them in a bunch, and let each think of his word as intently as he will; nowhere will there be a consciousness of the whole sentence.*
>
> *James (1890), Ch. VI*

As trivial as it may seem, James' thought experiment highlights the irreducible privacy of the mind. Each of the 12 men is aware only of his own word, and he is not aware of the others' words. Let us imagine now that we take 12 men and pinprick each of them. They will not be directly aware of what the others feel. Consequently, they will not be able to compare their pain with the others' pain and they will not be entitled to assume that they all share the same sensation. The problem of other minds arises from the privacy of the mind. It is important, however, to distinguish different versions of the privacy claim (Ayer 1963). Mental states are not private in the sense that one would be the only one able to *detect* one's own states. The men know that they are all in pain. The privacy of the mind does not entail the logical impossibility of mindreading. Nonetheless, the men do not detect the pain sensation in others *in the same way* that they detect pain in themselves. Mental states are private in the sense that one has a direct access to one's own mental states that nobody else has. There is an asymmetry between self-knowledge and knowledge of other minds.

However, Ayer (1963) challenges the hypothesis of the privacy of the mind by showing the conceivability of what he calls "co-consciousness", defined as follows:

> being able to make reports about the mental states of others in the same 'immediate' way as one makes reports about one's own, to report on them, in short, as if they were one's own (Ayer 1963, p. 65).

If indeed co-consciousness is possible, then mental states are no longer private and the problem of other minds is no longer a problem. One might even argue that

F. de Vignemont
Institut Jean-Nicod, EHESS – ENS – CNRS, 29, rue d'Ulm, 75005, Paris, France

co-consciousness is not only logically possible, but also empirically possible (de Vignemont 2004). Intersubjective phenomena such as empathy and motor resonance have indeed been recently interpreted in terms of "direct experiential understanding" of others (Gallese et al. 2004, p. 396). But do they really provide an access to other people's states "as if they were one's own"?

First-Person Perspective

Fifteen years ago, it was shown that performing an action and observing it activate the same internal representations of action, both in monkeys and in humans (Gallese et al. 1996; Rizzolatti et al. 1995; Grezes and Decety 2001). Shared cortical networks have also been found for empathy. Brain imaging studies have shown overlapping brain activation patterns when subjects feel pain and when they observe the others in pain (Singer et al. 2004), when they feel being touched and when they see someone else being touched (Keysers et al. 2004), when they inhale disgusting odorants and when they observe disgust-expressive faces (Wicker et al. 2003). These results are confirmed by lesion studies. For instance, NK is impaired both in the experience of disgust (as contrasted with fear and anger) and in disgust recognition (Calder et al. 2000). By contrast, SM is impaired both in the experience of fear and in fear recognition (Adolphs et al. 1994).

By linking self and others within a unique representational framework, these so-called mirror systems were soon considered at the core of intersubjectivity (Gallese 2003). The same representations are used both for self and for others, and as such, might be qualified as 'shared' (de Vignemont 2004; Jeannerod and Pacherie 2004). One consequence of the existence of shared representations is that they must be neutral relative to the agent/subject. They do not represent who is acting, just the type of action.[1] They do not represent who is feeling, just the type of emotion. Another possible consequence is that they enable the observer to recognize the intention, the emotion and the sensation displayed by other people. One might even suggest that one can directly perceive someone else's intentions by activating a similar intention in oneself (Iacoboni et al. 2005) and directly grasp what another feels by sharing the same bodily sensation or the same emotion (Gallese 2001). The general hypothesis can be articulated as follows:

1. Two tokens of the same type of representation are activated in two individuals' brains.
2. The activation in one individual is triggered by the other individual's state.
3. The activation of the same type of representation in the two individuals enables the observer to categorize what the other is doing/feeling.

---

[1]The content of shared representations of action can be described as follows: <x, action, object>, x as the unfulfilled parameter of the agent.

This description, however, does not fully capture the specificity of mirror systems. It can indeed apply as well to any kind of concepts. Concepts can be applied both to the self and to others. It is enough to master semantic knowledge about emotions or actions to apply it independently of who is feeling or moving. However, semantic knowledge is from a third-person perspective only. It does not provide any guarantee that one is able to feel or do the same from a first-person perspective. For instance, a ballet critic can watch a dancer and recognize the dance steps. The dancer and the critic share the same semantic representation of the dance steps from a third-person perspective. The ballet critic is, however, completely unable to dance. He merely conceptually grasps what the dancer is doing.

If the sharing were limited to conceptual recognition, there would be no particular interest in motor resonance and empathy. However, it is not because one can categorize the other's intentions on the basis of action observation that one should reduce mirroring processes to categorization. This is merely one functional implication of mirroring processes among others (e.g., imitation); this is not its definition. It is therefore important to dissociate the mirroring process and its possible use by the recognition process (Goldman 2006). Mirroring can serve as the basis of judgments about other people, but mindreading is not constitutive of mirroring. What makes motor resonance and empathy special is that they go beyond mere conceptual sharing. One does not share only the knowledge about the action or the emotion; one shares the subject's inner perspective. Mirroring allows the observer to internalize someone else's actions and emotions as if she were the subject of these experiences, and not just an external witness. In this sense, the mirroring process involves shared representations from a *first-person perspective*.

Several results argue in favour of this view. It was found that some patients may have a preserved declarative knowledge about a specific emotion, and yet be unable to recognize the emotion in others. This impairment is correlated with their inability to experience the specific emotion. For instance, SM who is impaired in fear perception knows what fear is and what kinds of situation trigger fear, but does not feel fear nor does he recognize fearful facial expressions (Adolphs et al. 1994). In addition, Blakemore et al. (2005) described a subject for whom the observation of another person being touched not only activated the mirror system for touch, but also induced a conscious experience of tactile stimulation on the equivalent part of her own body. This synaesthetic touch felt like real touch (Banissy and Ward 2007). What is shared is shared from the first-person perspective of the experiencing subject.

Similarly, for action, what is shared is shared from the first-person perspective of the agent in interaction with the world. A recent brain imaging study has compared participants with visual familiarity (third-person perspective) and participants with motor familiarity (first-person perspective) of the same observed action (Calvo-Merino et al. 2006). Ballet dancers, both male and female, watched short videos of ballet movements, either exclusively performed by male dancers, exclusively performed by female dancers, or common to male and female. The underlying assumption was that dancers have visual familiarity for all the movements, but only motor familiarity for movements of their own gender. If there were nothing

more to shared representations of action than a mere semantic representation of dance, then there should be no difference between the conditions. However, this was not the case: mirror system activity was greater when subjects watched movements specific to their own gender. Questionnaire data and brain activity suggested that semantic knowledge of the different movement types, such as visual recognition and naming, did not differ between the groups. This study provides a conclusive argument in favour of the existence of representations of action shared between self and others from a first-person perspective. They encode the properties of the motor system for optimization of action execution, as well as the properties of the object relevant for the interaction between the agent and the object. They allow someone else to reproduce the same movement through imitation.

We have seen conclusive empirical results in favour of representations from a first-person perspective shared between self and others. However, one may find such representations conceptually impossible. Some may indeed assume that the first-person perspective is necessarily self-specific. If the representation is from a first-person viewpoint, it must represent the agent of the actions and the subject of the emotional experiences as being oneself. Consequently, a representation could not be both from a first-person perspective and neutral relative to who the agent/subject is.

This criticism, however, is based on a fundamental confusion. One must distinguish between the descriptive content of the representation and its mode of presentation. For instance, the descriptive content of an action may be something like <I grasp the glass>. It includes the agent, the type of the action and the goal. By contrast, the mode of presentation constitutes the way the action is represented. The same action can be presented under the mode M 'from the inside' and under the mode M' 'from the outside', and they are not intensionally transparent. Let us go back to the ballet critic. He knows under the mode M' what an entrechat is, but for all that he does not know under the mode M what it is like to make an entrechat, in the same way that one can know which star is Hesperus without knowing which star is Phosphorus. The agent component of the content and the mode of presentation are therefore two separate dimensions. There is no contradiction in representing an action both with an unfulfilled parameter for the agent and from the first-person perspective (M mode).

To sum up, motor resonance and empathy rely on representations of the relevant motor intention, emotion or sensation, which are shared between self and other. In contrast with semantic representations, these shared representations are of particular interest as their content is presented from the inside, from the first-person perspective of the agent and the experiencing subject. This argues in favour of the empirical possibility of co-consciousness. A first requirement is fulfilled: one has access to other people's mental states under the same mode of presentation that one has access to one's own states.

However, it is not enough for the mode of presentation to be the same, the descriptive content must also be the same. This is indeed at the core of the shared representations view: two tokens with the same content (condition (1)). Only the similarity of content can guarantee the reliability of mindreading. But to what

extent are shared representations really identical? Does one have access to the full-fleshed states of others, or merely to a rough-grained copy of them?[2] In addition, this access must be direct. This claim is recurrent in the literature about empathy and motor resonance, which are said to allow: "to directly understand the meaning of the actions and emotions of others by internally replicating ('simulating') them without any explicit reflective mediation." Gallese et al. (2004), p. 396). They are supposed to provide a *direct* access to others in the same way that we have a direct access to ourselves. But is it really so?

If co-consciousness is more than a mere logically conceivable thought experiment, one should be able to show that empathy and motor resonance meet the two conditions of isomorphism and immediacy. However, as I will argue in the next two sections, this is true, neither for empathy, nor for motor resonance. They may represent other people's mental states from a first-person perspective; yet, the knowledge of others that they provide is still far from self-knowledge.

The Limits of Empathy

> For we certainly believe ourselves to be directly acquainted with another person's joy in his laughter, with his sorrow and pain in his tears, with his shame in his blushing, with his entreaty in his outstretched hands, with his love in his look of affection, with his rage in the gnashing of his teeth, with his threats in the clenching of his fist, and with the tenor of his thoughts in the sound of his words. (Scheler 1923, p. 254)

Scheler defended the view that emotional *sharing* enables us to *directly* understand someone else's emotions. More recent definitions of empathy assume that there is empathy if and only if the observer feels the other's emotion (Snow 2000)[3] and that the observer's emotion is activated automatically and directly by the observation of the other's emotion (Preston and de Waal 2002; Gallese 2001). Empathy is thus supposed to meet the conditions of isomorphism and immediacy, and as such, constitutes a valid candidate for co-consciousness. However, this analysis faces several difficulties due to a fundamental ambiguity in the literature about what counts as empathy. I suggest here distinguishing between mirror empathy and reconstructive empathy (de Vignemont 2008; Goldman 2008). The problem is that only the latter meets the condition of isomorphism, while only the former meets the condition of immediacy (or quasi-immediacy as I will argue). Consequently, there is no type of empathy that fulfils both requirements at the same time.

---

[2]This question is made even more salient by the fact that brain imaging studies reveal only partial overlap between execution/experience and observation. It is not everything that is shared.

[3]This is indeed what makes the difference between empathy and sympathy (de Vignemont and Singer, 2006). Empathy and sympathy are often taken as synonymous. However, in sympathy, one does not feel the same type of emotion as the other. For instance, I feel sorry for you because you feel jealous, depressed or angry but I am not jealous, depressed or angry myself, unlike in empathy.

Reconstructive and Mirror Empathy

Let us compare the two following quotations, both describing empathy:

> Attended perception of the object's state automatically activates the subject's representations of the state, situation, and object, and that activation of these representations automatically primes or generates the associated autonomic and somatic responses, unless inhibited. (Preston and de Waal 2002, p. 4)
>
> It [empathy] necessarily involves bringing to bear in the imaginative process a characterisation of the narrator, which will include facts about the narrator – not just aspects of his character, his emotions, moods, and other dispositional psychological facts about him (being irritable, loving his wife, having a phobia about dogs, being depressed), but also other not obviously psychological facts about him (being short, being brought up in 1960's Alabama). (Goldie 1999, p. 411)

Preston and Waal on the one hand and Goldie on the other hand agree on the definition of empathy as emotional sharing. However, they disagree on the underlying mechanism leading to emotional sharing. They appeal to two different theoretical frameworks. Preston and Waal interpret empathy in terms of the Perception-Action Model, like motor resonance, whereas Goldie interprets empathy in terms of imagination based on rich prior knowledge of the person one empathizes with. These two frameworks are, however, not incompatible. They describe two different routes to empathy, which I call mirror empathy and reconstructive empathy. Both routes towards empathy involve emotional sharing. But emotional sharing is generated by different mechanisms. To illustrate this distinction, let me describe two situations.

1. I see a stranger in the train smiling. It makes me smile. I recognize that she is happy, although I know nothing about her. I feel happy with her.
2. I learn indirectly that a friend is pregnant. I know how much she wanted a child. I put myself in her shoes and I realize how happy she must be. I feel happy with her.

Mirror empathy is induced by *the mere observation of emotional cues*. It is based on very poor input. It is enough to see a smile or to hear the tone of the voice to activate the feeling of happiness or anger. One does not need to know why the person is smiling or upset.

By contrast, reconstructive empathy is induced by *the simulation of the emotional situation*. One tacitly takes the perspective of another person, pretending to have her mental states and to be in the same situation. The input is thus more complex and requires extracting and selecting the relevant information. The different kinds of information are combined and evaluated, which then generate emotional states similar to the person's states.

Both routes to empathy involve the activation of an emotional state isomorphic to the target's state. Both routes to empathy enable us to understand the other's emotional state. Can they both qualify as co-consciousness?

Condition of Isomorphism

In order to provide a more accurate analysis of the notion of shared representation of emotions in empathy, it is useful to distinguish four main dimensions of emotional states: the type of emotional state (emotional valence: negative versus positive; emotional category: anger, happiness, disgust, fear, etc.), its focus (evaluation or appraisal of a situation that leads to the emotional state), its functional role (action tendency and emotional response), and its phenomenology (vividness, intensity, duration). We said that both routes towards empathy involve feeling the same emotion. But must the observer's emotion be strictly similar on all these different aspects? When I see you crying out of despair, am I desperate like you or do I just feel bad? To which degree is the empathic response isomorphic to the original affective state? Is it a coarse-grained congruency or a more fine-grained equivalence?[4]

Let me start with mirror empathy. We have seen previously that impairments in emotion experience and recognition are emotion-specific. They selectively affect fear or disgust. Emotion observation thus activates the representation of the same *type* of emotion, enabling the observer to categorize it. However, the activation does not need to reach the same level of intensity and vividness. When I see the stranger in the train smiling, I am not necessarily as happy as the stranger. But this does not prevent me to recognize the observed emotion. I know it is a positive feeling, and even more, I can pinpoint that this is happiness.

What about the *focus* of the emotion? Do I feel happy for the same reasons as you? Quite often we see people with facial expressions indicating different types of emotions without knowing why they feel like that. One can hardly assume that we share the same focus. Rather, we feel happy for no good reason. If we had to explain our emotion, it would be difficult.[5] If we are limited to bodily cues, then we do not have access to the subjective meaning of the emotion (Schutz 1932). For instance, we cannot distinguish when the emotion is real and when it is fake. The perception of a happy face allows us to directly recognize and categorize the kind of emotion, but it does not allow us to understand its cause and its reason.

As for the *functional role*, it cannot be fully the same, for the reasons just explained. Let us imagine that your face reveals that you are in pain, but I cannot see what is hurting you. I have an empathic response, activating the affective

---

[4]Neuroscience cannot yet provide an unambiguous answer. Some empathy studies find only activity in the affective component of the pain network (Singer et al. 2004; Jackson et al. 2005), whereas others show reduced motor excitability specific to the muscle that the subjects observed being hurt (Avenanti et al. 2005). Likewise, a recent study of empathy for touch revealed activation of primary sensorimotor cortex that was somatotopically mapped (Blakemore et al. 2005), contrasting with another study showing only secondary somatosensory cortex activity (Keysers et al. 2004).

[5]This comes closer to emotional contagion. The only difference is that mirror empathy is directed toward other: we feel happy with the other and we are aware that the other is happy. For a discussion about the relationship between mirroring, emotional contagion, and empathy, see de Vignemont (2009).

component of my pain network, but my sensation of pain remains indefinite. It is not a headache or a toothache. It is just an indeterminate pain. It might increase some basic physiological functions (e.g., heartbeat) like any real pain, but it cannot activate the appropriate response tendency because of this indeterminacy (e.g., remove my hand from the hot pot). Consequently, in mirror empathy, the observer's sensation/emotion cannot play the same functional role as the target's sensation/emotion.[6]

We have seen that the isomorphism between the observer's emotional state and the target's state is limited to the type of emotion, and does not go beyond. It cannot provide a full-blown access to the other's emotional feelings, in the same way that we have a full-blown access to our own emotional states. Mirror empathy meets only partially the condition of isomorphism.

Emotional sharing may be more exhaustive in reconstructive empathy. One puts oneself in the other's shoes and runs one's own affective appraisal of the situation. For instance, I pretend that I am pregnant and that I want a child, which leads me to feel happy. My emotional state is about the pregnancy. It has the same focus as my friend's emotion. The causes and reasons of the emotional state are indeed the input of the simulative process. They are constitutive of reconstructive empathy. Furthermore, reconstructive empathy is of particular interest for mindreading because it provides access not only to the fine-grained emotional states, but also to the other mental states that are associated to them. The output of the simulative process is not exclusively emotional. It also includes the beliefs, desires and intentions that are triggered by the emotional situation. For instance, I may have the tendency to start imagining the name of the baby, although this reaction would probably be inhibited. My empathic response can thus play partially the same functional role as my friend's emotion. It allows me to predict my friend's behaviour, in contrast with mirror empathy. The isomorphism between the observer's emotional state and the target's state is thus more extended in reconstructive empathy, although it never reaches a perfect similarity.

Empathy provides access to other people's mental states from a first-person perspective. However, in mirror empathy, other people's mental states one has access to are restricted in scope and limited in content. They are very basic and do not afford a full grasp of the others, nor do they allow predicting their behaviour. Mirror empathy isolates the emotion from the rest of the mental life. It is not enough to share the same type of emotion for having access to other's emotions as if they were one's own. One needs a finer-grained sharing, based on a common focus and on a common network of associated mental states. This is provided by reconstructive empathy, which relies on a holistic approach on the other person's emotional life. It is only in reconstructive empathy that knowledge of other people gets close to self-knowledge because it embraces the whole emotional life of others, but there is a cost to pay for a deeper understanding of others, as we will see now.

---

[6] Here I focus on affective mirroring, which is the most common during pain observation (Singer et al. 2004; Jackson et al. 2005), rather than on sensorimotor mirroring (Avenanti et al. 2005), which is closer to emotional contagion than to empathy.

Condition of Immediacy

I will start this time with reconstructive empathy, which appears to be the most promising candidate for co-consciousness. It does indeed meet the condition of isomorphism, but does it also meet the condition of immediacy? Does one have a *direct* access to the other's full-blown emotional states?

It depends on what the condition of immediacy involves. According to one interpretation, one may argue that reconstructive empathy fulfils this criterion. The argument goes as follows. In reconstructive empathy, one has a direct access to one's own emotion. One's own emotion is similar to the other's emotion. Thus, one has a direct access to the other's emotion. If this argument is valid, it could be extended to all mental states that one has access through simulation. There would be no more asymmetry between self-knowledge and knowledge of others. However, the proponents of the Simulation theory have never pretended to erase the asymmetry. On the contrary, self-knowledge is based on introspection, whereas knowledge of others is based on simulation (Goldman 2006). Furthermore, it sounds counterintuitive to claim that it is enough to pretend to have someone else's states to have a direct access to her states. There is a direct access if and only if the causal transmission of information is direct and does not involve intermediary steps. To have a direct access to the other's states, I need information not only about my pretended states, but also about the causal relationship between the other's states and my states. Although simulation may involve introspection at the end in order to determine the output of the simulative process, it also involves other processes such as selection of information, pretence, decision-making mechanism, appraisal mechanism, etc. (Goldman 2006). We are far from a one-step process. In this sense, reconstructive empathy does not meet the condition of immediacy because it does not provide a direct access to *the other's emotion*, but merely to *one's own emotion*, which happens to be similar to the other's emotion.

What about mirror empathy? In contrast with reconstructive empathy, it arises from the mere perception of the other person. According to Scheler, one is 'directly acquainted' with the other's emotion. There would be no need for inferences from your facial expression to your internal emotion. The link between them would be so direct that I would be able to automatically 'read' your emotion (Preston and de Waal 2002). Put it another way, the emotional cue and the emotion itself would be one and the same thing. Consequently, mirror empathy would be based on a univocal function that goes straightforwardly from the cue to the internal state. It should not take into account any other parameter from the context. Unlike reconstructive empathy, which is based on a rich set of information, mirror empathy is based on very restricted inputs, and it is only this poverty of stimulus that can guarantee its immediacy.

However, I would like to argue that even mirror empathy is not that simple. It is not true that each time I perceive an emotional cue, I immediately and automatically have the appropriate empathic response. Even in mirror empathy, the empathic response is mediated by the context, as shown by the following examples.

(a) *Multiple cues*: What happens when you are confronted to several persons in different emotional states, each of them displaying different emotional cues? According to the view just described, you should 'directly' read each emotion by having an empathic response for each emotion. However, that may lead you in counterintuitive situations. Imagine that you see a mother yelling at her child. She looks very upset and he looks very sad. Do you empathize with both of them? One may reply that the activation of one of the emotions inhibits the activation of the other. But why this emotion rather than that one? Do we choose with whom we empathize? Some perceptual factors may of course intervene, like the saliency of the emotion. However, it is highly plausible that there are other external factors that are not purely perceptual. For instance, at the beginning I may feel more empathy for the child, but if I learn that the mother is upset because the child hurt his little brother, then I will probably feel more empathy for the mother. The empathic emotion that I feel will thus be determined by contextual factors, and as such, it is not direct.

(b) *Ambiguity*: One may still argue that in this case, you had first an empathic response for both emotions, but that it was later on inhibited by further information. However, this cannot account for all cases. Sometimes perception of facial expression does not suffice in itself and you need beforehand other types of information in order to categorize the emotion. For instance, if I see you blushing, Scheler claims that I am directly acquainted with your shame. However, how can I know that you are blushing because you feel shameful or because someone complimented you and you feel happy? On the only basis of the colour of your cheeks I cannot take side, and even less feel empathy. I need further information before being able to share your emotion. I am not 'directly acquainted' with your shame.

(c) *The relationship with the target*: Nonetheless, one may maintain that these examples are just exceptions, and that most of the time we have a straightforward access to the other's emotion. How to deny that if you see someone in pain, you automatically feel pain with her? However, it is not just anyone with whom we empathize; it is a person with whom you may have a specific relationship. And this relationship is also a parameter to take into account and it may affect your empathic response. Imagine that you are playing at a money game with people that you just met tonight. One of them is unfair and treats you badly. Just after the game, he hurts himself. Do you feel empathy for him, although you do not like him? Or rather do you feel that he just got what he deserved? This last scenario is actually the design of an fMRI experiment (Singer et al. 2006). The women participants still showed the activation of the pain network. In contrast, the men participants showed none. Consequently, the relationship that one has with the target modulates the empathic response. Other contextual parameters can also affect it, like the reason of your emotion. Could I share your joy if I knew that it was not justified? It was recently shown that subjects have less empathetic response in pain-related areas when they know that the pain inflicted to the other was useful (to cure the other), than when they know it did not help the person (Lamm et al. 2007).

It would be too long here to list all the different parameters that may modulate mirror empathy. What matters is that in several cases mirror empathy is mediated by the processing of supplementary information. Even mirror empathy is not so low-level. It is not activated directly and automatically by the perception of an emotional cue. Rather, the emotional cue makes sense in the emotional context. Outside experimental situations, it is very rare that one perceives a facial expression without perceiving also other types of information about the person and her situation. Whether an empathic response is elicited depends on the outcome of the contextual appraisal process (de Vignemont and Singer 2006). The access to the other's emotion is not as direct as Scheler and others seem to assume. It is mediated by the analysis of the emotional context, which strongly influences the magnitude of empathic responses.

Is there still a difference between the mirror route and the reconstructive route to empathy, if both are context-dependent? I would like to keep the distinction. Mirror empathy does not require knowing everything about the emotional situation, in contrast with reconstructive empathy. It does not require pretending to have someone else's states. It is still more direct, because contextual modulation can be fast and implicit. In this sense, mirror empathy is quasi-direct.

Empathy has thus to face an alternative. Either it gives access to the full-fleshed emotional life of others. But it relies on a series of basic processing like information selection, pretence and affective appraisal. We are far from the immediacy of self-knowledge. Or it gives a quasi-direct access to the emotions of others, although modulated by contextual appraisal. But it provides a very restricted knowledge limited to the kind of emotion. We are far from the richness of self-knowledge. The two conditions for co-consciousness cannot be met both at the same time.

The Limits of Motor Resonance

> Action observation causes in the observer the automatic activation of the *same* neural mechanism triggered by action execution. It has been proposed that this mechanism could be at the basis of a *direct* form of action understanding. (Gallese 2007, p. 6, my underline)

The literature about motor resonance is more recent than the empathy literature, but these last 15 years it has known an increasing interest both on the empirical and on the philosophical side. There are two main reasons for that. First, as said in the first section, it is supposed to rely on action representations shared between self and other (condition of isomorphism). Second, it is supposed to provide a direct perception of other people's motor intentions (condition of immediacy). However, as we will see, motor resonance faces difficulties similar to those encountered in mirror empathy.

Mirror systems are defined as the brain areas that are activated during both action execution and action observation, but these last two notions are too broad to be useful to understand what exactly is common. Both cognitive and neural models

identify many representational stages and processes that contribute to action execution (Jeannerod 1997; Wolpert et al. 1995). Exactly what information is shared? When I see you grasping the doorknob, do I have access to your intention to open the door? To go out? To buy some bread? And is this access direct or mediated by contextual appraisal? Like for empathy, we will see that the deeper our understanding of others is, the less direct it is.

Condition of Isomorphism

> What if the man wielding the axe is not really chopping wood but merely appears to be doing so? What if the man holding the doorknob is not grasping it in order to shut the door, but is merely holding it steady in order to repair it? What if the hunter is not taking aim at all but is merely watching the animal through the telescopic sight of his rifle? Observational understanding of the other person's outward behaviour is clearly not enough to settle these points. Schutz (1932, p. 27)

Schutz highlights the limits of mindreading based on action observation. Observing bodily movements does not provide a full-blown access to the other's intentions and to the subjective meaning of the action. When I observe someone performing a movement, it activates the representation of the movement, but it does not provide a full understanding of the observed movement. The point raised by Schutz is that the mere perception of movements leaves open a number of possible interpretations. I have argued that in mirror empathy the isomorphism between the observer's emotional state and the target's state is limited. Here, I will tackle the problem in a slightly different way. The peculiarity of the motor system – in contrast with the emotional system – is that it is hierarchically organized. The goal of the action is represented at different levels of complexity (i.e. a set of motor commands, a series of intermediary movements, the final outcome) (Wolpert et al. 2003). As Anscombe (1957) said, there is an almost infinite list of possible descriptions of the same action. To analyze the isomorphism between the observer's motor representation and the target's representation, one needs to take into account the complexity of the motor system to determine the level that is shared within the hierarchical structure of the motor system.

At the computational level, hierarchical models of motor control postulate the existence of at least three layers of motor representations, that one may call the prior intention (i.e., the goal), the intention in action (i.e., the sequence of movements) and the motor command (e.g., the elements of movements) (Searle 1983; Wolpert et al. 2003). The *prior intention* is the highest and the most abstract level. It represents the goal of the action as a global unit (e.g. to drink). However, there may be several means to realize the same prior intention. The motor system needs to anchor the prior intention in a specific context and to determine the dynamic sequence of specific movements that will be required to reach the goal, that is, the *intention in action* (e.g. to pour the water in the glass, to grasp the glass and to bring it to the mouth). The *motor command* takes into account the agent's body and the surrounding context and represents the precise kinematics of the sequence of

movements (e.g. to stretch the arm for 20 cm, to make a 5 cm grip aperture, to raise the arm with an elbow angle of 45°).

Most of the neuroscience literature about motor resonance does not distinguish clearly among the different levels of the motor hierarchy. They tend to oscillate between two extremes. On the one hand, mirror neurons were said to be a replicative mechanism based on a one-to-one matching. They directly duplicate in a motor code the perceptual properties of the observed movement. On this view, they must be fine-grained representations of action, like motor commands. On the other hand, mirror neurons were said to represent the agent's global intention. The observation of the same sensory outcome can indeed lead to different activation of mirror systems depending on the intention (Fogassi et al. 2005; Iacoboni et al. 2005). Mirror neurons must then be more abstract and rough-grained representations of action, like prior intentions.

It seems that the truth is in between (de Vignemont and Haggard 2008). On the one hand mirror systems do not represent the bodily kinematics of the observed movements in all their details. What matters is that the observer can perform a functionally equivalent sequence of movements, not the way this sequence is implemented in the agent's particular body. Not only do mirror neurons match the observed action to an executable action, but also someone else's body to one's own body. On the other hand they are not very abstract representations of the prior intention. As shown by Schutz and more recently by Jacob and Jeannerod (2005), there is not enough information in the perceived actions to actually understand the other's prior intentions: if you see a man grasping a scalpel and applying it to someone's body, you cannot know, on the only basis of the visual information, whether it is to save her life or to kill her, whether the man is Dr Jekyll or Mr Hyde. To empirically show that mirror systems detect prior intentions, one would need to find similar brain activations when the displayed movements are completely different, although the global goal remains the same. And as far as I know, this is not the case.

What is shared is thus the intention in action, which represents (i) bodily movements, independently of the agent's or the observer's specific bodily parameters to allow transfer across individuals in imitation, (ii) at the intermediate level of the planning of the dynamic sequence of movements. Because the isomorphism is limited at the level of intentions in action, the knowledge that one can get on the basis of motor resonance is restricted. On the one hand, it is restricted because it does not give access to all the motor details. On the other hand, it is restricted because it does not give access to the abstract level of other people's prior intentions. One is far from the knowledge of the other people's will.

Condition of Immediacy

Nonetheless, one may maintain that motor resonance instantiates a kind of minimal co-consciousness. It is indeed supposed to provide a *direct* access to others in the same way that we have a direct access to ourselves. As we have seen with mirror

empathy, direct understanding is possible if and only if there is a one-to-one relationship between the perceived input and its meaning, when there is no room for interpretation. The Direct Matching hypothesis has emphasized the congruence between the perceptual and the motor properties of the action: a neuron that is specifically tuned for execution of a specific motor action shows visual selectivity for observing the same action (Rizzolatti et al. 2000; Gallese 2003). Mirror neurons were said to be merely a duplicative mechanism.

However, some results in the mirror system literature do not fit with the Direct Matching hypothesis (for review, see Csibra 2007; Jacob 2008). First, one may otice that the similarity between the perceptual input and the representation of action can be very dim. It was shown indeed that some mirror neurons responded to the observation of tool-using actions (e.g. grasping food with a stick), although the monkeys themselves were unable to perform these actions (Ferrari et al. 2005). In this case, the kinematics of the two actions differs in important respects (e.g. the direction and speed of motion, the angles between the joints, etc). More generally, mirror neurons are far from being a faithful reproduction of what is observed. Only 19% to 41% of mirror neurons are classified as 'strictly congruent'. Most of them fire responding to movements as different as grasping with the mouth and grasping with the hand (Gallese et al. 1996). Different perceptual inputs can thus activate the same motor representation. Conversely, the same perceptual input can give rise to different motor activations depending on the context (Fogassi et al. 2005; Iacoboni et al. 2005). Iacoboni et al. (2005) found different activations of the human mirror system depending on the intention guiding the same movement (e.g. grasping the cup for drinking or for cleaning). However, if the action was presented without any context, there was no difference. What matters was that the observed actions were embedded in different contexts (before tea or after tea).

Csibra (2007) concludes that action understanding is a prior necessary condition for mirror activation, rather than a consequence. What is mirrored is not an uninterpreted visual signal. There is a visual decomposition of the observed action into relevant units, disregarding some aspects of it (e.g., the position of the non-active hand in grasping) as irrelevant for reproduction. To evaluate what is relevant or not to map, one needs to interpret the bodily information within its context to reconstruct its goal. Action understanding arises from the integration of different sources of information. There is no direct matching. Mirror activations are mediated by the interpretation of the perceived movement.

To illustrate the debate, I will contrast two imitation studies. In the original Meltzoff's experiment, children see the experimenter touching with his forehead a box that makes sounds. Meltzoff (1988) found that they imitate the action, replicating faithfully the head-touch movement. However, 15 years later, Gergely et al. (2002) used the same set-up, but with two different conditions. In one condition, the experimenter had his hands free, while in the other, he had his hands occupied. The authors found that infants imitated the head-touch action only in the first condition. What is relevant to imitate is thus evaluated in terms of the global context. The choice of the effector depends on what is available to the agent. In the second condition, the experimenter had his hands occupied, which explains why he used his head,

but this did not apply to the infants who had their own hands free, and who could use them. Imitation is not a mere copy of what is seen based on the direct access to the other's intention. It relies on the contextualized interpretation of the intention in action and the reconstruction of the action to perform.

As in mirror empathy, motor resonance seems at first sight a simple mechanism of understanding other people that relies on direct mapping of perceptual representations of action unto motor representations. However, the relationship between the observed action and the mirror state is more complex. It is not encapsulated, isolated from other contextual information. On the contrary, it takes into account external factors, based on the visual analysis (e.g., is the cup full or empty?) and rational analysis (e.g., are the hands free or occupied?) of the scene. Consequently, the same perceptual input can activate different motor representations and conversely, different perceptual inputs can activate the same motor representation. We are far from the one-to-one mapping that would enable a direct access to the other.

Conclusion

According to Schutz, it is not enough to understand the objective meaning of the behaviour, one must also understand what the behaviour means for the other person and why she is using it here and now. This cannot be provided by mirror empathy, nor by motor resonance. All one can do is merely categorize and recognize the other's action and emotion, not to understand them. Even if we are far from the richness of access that we have to our own mind, with all its turmoil of complex emotions and intentions, one may argue that we still have access to the other's mind as if it were one's own, except that it would be a very poor mind, limited to very basic states with simple content. True, mirror empathy and motor mirroring are not far from providing a direct access to the others. Yet, they do not fully meet the condition of immediacy. One should not neglect the intermediary stages of contextual appraisal that modulate both mirror empathy and motor resonance. To conclude, Ayer's notion of co-consciousness remains logically conceivable, but it has no empirical counterpart. This is not to deny the interest of intersubjective mirroring, which allows us to internalize other people's states from the first-person perspective.

References

Adolphs R, Tranel D, Damasio H, Damasio A (1994) Impaired recognition of emotion in facial expressions following bilateral damage to the amygdale. Nature 372:669–672

Anscombe GEM (1957) Intention. Harvard University Press, London

Avenanti A, Bueti D, Galati G, Aglioti SM (2005) Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. Nat Neurosci 8(7):955–960

Ayer AJ (1963) Privacy. In: The concept of a person and other essays. St Martin's Press, New York, pp 52–81

Banissy MJ, Ward J (2007) Mirror-touch synaesthesia is linked with empathy. Nat Neurosci 10:815–816

Blakemore S-J, Bristow D, Bird G, Frith C, Ward J (2005) Somatosensory activations during the observation of touch and a case of vision-touch synaesthesia. Brain 128:1571–1583

Calder AJ, Keane J, Manes F, Antoun N, Young AW (2000) Impaired recognition and experience of disgust following brain injury. Nat Rev Neurosci 3:1077–1078

Calvo-Merino B, Grezes J, Glaser DE, Passingham RE, Haggard P (2006) Seeing or doing? Influence of visual and motor familiarity in action observation. Curr Biol 16(19):1905–1910

Csibra G (2007) Action mirroring and action interpretation: An alternative account. In Haggard P, Rosetti Y, Kawato M (eds) Sensorimotor foundations of higher cognition. Attention and performance XXII. Oxford University Press, Oxford

de Vignemont F (2004) The co-consciousness hypothesis. Phenomenol Cogn Sci 3:97–114

de Vignemont F (2008) Empathie miroir et empathie reconstructive. Revue philosophique de la France et de l'Etranger 133(3):337–345

de Vignemont F (2009) Affective mirroring: emotional contagion or empathy? In: Atkinson and Hilgard's introduction to psychology

de Vignemont F, Fourneret P (2004) The sense of agency: a philosophical and empirical review of the "Who" system. Conscious Cogn 13(1):1–19

de Vignemont F, Haggard P (2008) Action observation and execution: what is shared? Soc Neurosci 3(3):421–433

de Vignemont F, Singer T (2006) The empathic brain: how, when and why. Trend Cogn Sci 10(10):435–441

Ferrari PF, Rozzi S, Fogassi L (2005) Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex. J Cogn Neurosci 17:212–226

Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, Rizzolatti G (2005) Parietal lobe: from action organisation to intention understanding. Science 308:662–667

Gallese V (2001) The 'Shared Manifold' hypothesis: from mirror neurons to empathy. J Conscious Stud 8(5–7):33–50

Gallese V (2003) The manifold nature of interpersonal relations: the quest for a common mechanism. In: Frith C, Wolpert D (eds) The neuroscience of social interaction. Oxford University Press, Oxford

Gallese V (2007) Embodied simulation: from mirror neuron systems to interpersonal relations. Novartis Found Symp 278:3–12

Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996) Action recognition in the premotor cortex. Brain 119:593–609

Gallese V, Keysers C, Rizzolatti G (2004) A unifying view of the basis of social cognition. Trend Cogn Sci 8(9):396–403

Gergely G, Bekkering H, Kiraly I (2002) Rational imitation in preverbal infants. Nature 415(6873):6755

Goldie P (1999) How we think of others' emotions. Mind Lang 14(4):394–423

Goldman AI (1993) Consciousness, folk psychology and cognitive science. Conscious Cogn 2:364–382

Goldman AI (2006) Simulating minds: the philosophy. Psychology and neuroscience of mindreading. Oxford University Press, Oxford

Goldman AI (2008). Two routes to empathy: insights from cognitive neuroscience. In Coplan A, Goldie P (eds) Empathy: philosophical and psychological perspectives. Oxford University Press, Oxford

Grezes J, Decety J (2001) Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. Hum Brain Mapp 12(1):1–19

Iacoboni M, Molnar-Szakacs I, Gallese V, Buccino G, Mazziotta JC, Rizzolatti G (2005) Grasping the intentions of others with one's own mirror neuron system. PLoS Biol 3(3):529–535

Jackson PL, Meltzoff AN, Decety J (2005) How do we perceive the pain of others? A window into the neural processes involved in empathy. Neuroimage 24:771–779

Jacob (2008), What do mirror neurons contribute to human social cognition? Mind Lang 23(2):190–223

Jacob P, Jeannerod M (2005) The motor theory of social cognition: a critique. Trend Cogn Sci 9(1):21–25

James W (1890) Principles of psychology. Harvard University Press, Harvard

Jeannerod M (1997) The cognitive neuroscience of action. Blackwell, Oxford

Jeannerod M, Pacherie E (2004) Agency, simulation and self-identification. Mind Lang 19(2):113–146

Keysers C, Wicker B, Gazzola V, Anton J-L, Fogassi L, Gallese V (2004) A touching sight: SII/PV activation during the observation of touch. Neuron 42:335–346

Lamm C, Batson CD, Decety J (2007) The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. J Cogn Neurosci 19(1):42–58

Meltzoff AN (1988) Infant imitation and memory: nine-month-olds in immediate and deferred tests. Child Dev 59(1):217–225

Preston SD, de Waal FBM (2002) Empathy: its ultimate and proximate bases. Behav Brain Sci 25:1–20

Rizzolatti G, Fadiga L, Gallese V, Fogassi L (1995) Premotor cortex and the recognition of motor actions. Cogn Brain Res 3:131–141

Rizzolatti G, Fogassi L, Gallese V (2000) Cortical mechanisms subserving object grasping and action recognition: a new view of the cortical motor functions. In: Gazzaniga M (ed) The new cognitive neuroscience. MIT Press, Cambridge, MA

Scheler M (1923/1954) The nature of sympathy. Yale University Press, New Haven, CT

Schutz A (1932/1967). The phenomenology of the social world. Northwestern University Press, Evanston, IL

Searle J (1983) Intentionality, an essay in the philosophy of mind. Cambridge University Press, Cambridge

Singer T, Seymour B, O'Doherty J, Kaube H, Dolan R, Frith C (2004) Empathy for pain involves the affective but not sensory components of pain. Science 303:1157–1162

Singer T, Seymour B, O'Doherty J, Stephan KE, Dolan RJ, Frith CD (2006) Empathic neural responses are modulated by the perceived fairness of others. Nature 439:466–469

Snow NE (2000) Empathy. Am Philos Quaterly 37(1):65–77

Wicker B, Keysers C, Plailly J, Royet J-P, Gallese V, Rizzolatti G (2003) Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. Neuron 40:655–664

Wolpert DM, Ghahramani Z, Jordan MI (1995) An internal model for sensorimotor integration. Science 269(5232):1880–1882

Wolpert D, Doya K, Kawato M (2003) A unifying computational framework for motor control and social interaction. In: Frith C, Wolpert D (eds) The Neuroscience of Social Interaction. Oxford University Press, Oxford

Intersubjectivity, Cognition, and Language

N. Praetorius

In our normal everyday encounters with other people it is generally assumed that the knowledge we have and the language we use to describe and communicate about things and events in the physical and social world of which we are part is *intersubjective*, that is, it is shared by the people with whom we may communicate and co-act. Indeed, this assumed intersubjectivity of cognition and language would seem to be a *precondition* for any co-action and linguistic communication to take place among people about things which exist in the so called "outer", publicly observable physical and social world. Arguably, it is a precondition for our very notion of a publicly observable physical and social world, i.e. a world that may be observed and described objectively and truthfully from a so called *third-person* view.

Whereas scepticism is no longer in fashion when it comes to the possibility of so-called *third-person* description and cognition of publicly observable reality being true or objective – indeed, not only our everyday encounters but also our scientific investigations rely on this possibility – there has for centuries been an extensive debate within philosophy, and more recently within the scientific studies of mind and consciousness, concerning the status of the description and cognition of things which do not exist in a publicly observable and describable world, such as our so called "internal" mental states. The uncertainty and scepticism are expressed in questions such as these: Given that mental states such as thoughts, emotions and feelings of pain, are not publicly observable, but may only be known or experienced by the persons who have them, how then can we be certain that these states exist and may be observed in the same way by different people? Furthermore, in view of the lack of public criteria or standards, how can we be certain that assertions put forward about such states have the same implications and use for different persons and language users? Or be certain that we are using language in the same way – or indeed using the *same* language – as we do when talking about and describing things in public material reality? Indeed, given that internal states are not the sort of things which are publicly *shareable*, how do we ever come to learn and talk about them in a language we do share?

N. Praetorius
Department of Psychology, University of Copenhagen

That solutions to these problems have serious consequences, not only for the possibility of scientific studies of mind and consciousness but for a science of psychology in general, is obvious when we consider that a crucial part of the knowledge a person has of himself, of his mind, body and acts – and, not the least, *of his mind and body being his*, and *of his acts being acts intended and controlled by himself* – rests on observations and experiences to which only the person himself has access. No one except the person himself will ever be able to experience what it is like to be the person he is *in the way he experiences* it, or to know what it is like to have his perceptions, thoughts, or feelings of pain in his body *in the way he does*, because no one except the person himself can see with his eyes, think his thoughts or feel his pain.

To this uniqueness of our personal experience must be added the differences in our knowledge and conception of ourselves, of others and the world in which we find ourselves, due, for example, to differences in our upbringing, training, education or cultural background. And yet, neither the knowledge of persons which is uniquely personal, nor the differences in knowledge and conceptions of themselves, of others and the world, due to their different backgrounds, can be said to be *private*. For, despite these differences in our personal experiences and conceptions, to be a person is something fundamentally social. Indeed, no one can be a person, and thereby someone who may realise that he or she is uniquely different from other persons, *without* other persons from whom he or she may differ – nor without having possibilities of determining *how* one differs from others – whether such differences concern one's notions and experiences of things in material reality, or one's inner feelings, thoughts, beliefs, emotions or pains. Indeed, it would seem that no one could be a person, that is, someone who *differs* and *knows* that one differs from others, without being able to communicate and talk with others about *how* one differs from them.

In what follows an attempt will be made to argue that conditions for intersubjective communication exist which apply equally for knowledge and description of things in publicly observable material reality and for our non-publicly observable mental or internal states. They are conditions on which rely the possibility of persons to communicate about their knowledge and experience, and conditions for determining and discussing their individual differences concerning their knowledge and experience – be it of that which is publicly observable or of that which is uniquely observable to themselves, such as their thoughts, emotions and feelings of pain. By the same token they are conditions on which relies the very possibility of persons developing together criteria or standards to determine and distinguish between what is publicly observable and what is personal, and by which, so it will be argued, any notion about private cognition and private languages are rendered untenable. Furthermore, it will be argued that despite the significant observational differences applying to the cognition and description of things existing in public material reality and to the cognition and description of our internal or mental states, we have to assume that the language in which we talk about both these things and states, is a language which necessarily relies on and is part of the very same language we use to describe things in public material reality.

The section which follows first presents the problems concerning knowledge and description of internal or mental states as set out within traditional philosophy. To this end an analysis will be carried out of the position on these problems defended by Wittgenstein. It is the aim of this analysis to show why it has to be assumed that, despite not being publicly observable, our knowledge and experience of such states are as amenable to intersubjective discourse as are things in the publicly observable world.

Conditions for Description of Mental or Internal States

One of the significant differences in the conditions for the cognition and description of things in publicly observable reality and of our non-publicly observable internal states concerns the *procedures* for determining the truth and correct application of descriptions of such things and states. In the case of descriptions of publicly observable things, for example, we will be able to take part in a procedure for determining whether the implications of the descriptions hold true for the things in question, and thus to observe the result of a test as to the correct application of the description. However, in the case when somebody says e.g. that he has a pain in his finger, no such public procedure exists for determining whether a "pain-description" is a true description of what the person feels in his finger. For only he knows and may observe what he feels in his finger.

But given that no public procedure exists for determining the correct use of our description of internal states such as feelings of pain, how then can we be sure that when different people talk about pains, they are talking about the same kind of "thing"? How, more precisely, *can I be sure that when I use the term 'pain', I use this term to refer to the same kind of "thing" as others do when they use the same term?* Or, conversely, *when other people use the term 'pain' do they then use this term to refer to the same kind of "thing" as I do when I use the same term?*

This would seem a perfectly sensible question. However, in his classical "Beetle in the box" example, in which Wittgenstein attempts to make clear what this question entails, it seems to be a question to which, for obvious reasons, no sensible answers may be given. Thus, Wittgenstein writes:

> If I say of myself that it is only from my own case that I know what the word "pain" means must I not say the same of other people too? And how can I generalise the one case so irresponsibly?

> Now someone tells me that he knows what pain is only from his own case!—Suppose everyone had a box with something in it; we call it a "beetle". No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at his beetle. Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. But suppose the word "beetle" had a use in these people's language? If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as something: for the box might even be empty. No one can 'divide through' by the thing in the box; it cancels out, whatever it is. (Wittgenstein 1945/1953, para. 293, p. 100.)

In the analysis which follows an attempt will be made to shown that the question about 'pain', posed in this *radical* sense, is not only a question which cannot be answered, but more importantly, it is also a question which cannot be *asked*. For it is a question in which the very condition for putting it forward is itself questioned – thereby debarring it of any sensible meaning.

Let me begin by noting that Wittgenstein does not only *suppose* that the word pain "has a use in people's language" – he knows it for a fact. Indeed, when he or anyone else asks the question, "do I use the term 'pain' to refer to the same "thing" as others do when they use the term", he and they are obviously asking this question in a language of which the term 'pain' is part – and thus is a term which is supposed to have a meaning and use which he and other people with whom he shares the language know *in virtue* of being speakers of that language, just as they know to what the term may be correctly applied, i.e. some particular unpleasant sensations somewhere in one's body. If Wittgenstein did not know this – and did not *presuppose* that all other speakers of the language knew this – neither he, nor they would have any idea what he is asking about, and no further discussion of the question would seem possible.

Now, it has to be admitted that we may have all kinds of difficulties in giving adequate accounts in words of the pain we may feel, and that we often feel uncertain about the choice of appropriate terms. Is this pain, for example, a sharp, piercing, dull, shooting, tender, searing pain, or is it a nagging or stabbing pain? Indeed, we may have great difficulties in conveying precisely to others the suffering we endure when we are in pain. However, these difficulties of *adequately* describing pains are not relevant to, nor do they invalidate the point just made. Indeed, these difficulties of adequate descriptions of pains, and discussions about them, could not take place *unless* people discussing them had a concept of and a term for pain which they shared, and knew of what it may be used correctly to refer to.[1]

We may contend then that Wittgenstein's discussion of the "language game" of giving expressions to pain relies on this knowledge and these presuppositions, and hence on the assumption that sensations and feelings of pain are the sort of "things", which do indeed exist and which we may use language to refer to. Furthermore, it

---

[1]To spell it out, behind questions such as "do we mean the same thing when we talk about 'pain' – or, for that matter about other mental or internal states such as 'memory' or 'recognition' – lies the assumption that such terms are no more randomly applied to mental or internal states of human beings than 'oak' and 'ash' are randomly applied to trees. Indeed, to take such questions seriously is already, albeit implicitly, to endorse certain assumptions of *how language functions in communicative contexts*. Thus, it is implied and assumed that it is possible for language users together to identify – if only rudimentarily – what they are talking about (e.g. some particular states or properties of our mind, or some sensations felt somewhere in our body), and that, on the basis of such common consent, it is possible to investigate whether the implications of those terms are in fact the same for everyone – and whether we in fact use these terms to refer to the same sort of "things". Conversely, such questions cannot be asked in any sensible way, nor may these terms be "mentioned" without or independently of how language and its terms are used to refer to actual things or events.

relies on the assumption that other people would use the term 'pain' to refer to the feeling of pain that he may have in his body – could they feel what he feels – and hence that he may indeed generalise from himself to others, when they use the term 'pain' – just as they may generalise from their use of the term to his and the use of the term by others.

However, contrary to the presuppositions on which his whole discussion of the question of 'pain' rely, Wittgenstein suggests, for the sake of argument, that because his feeling of pain, and those of others, are not publicly observable, his feelings of pain may be completely different from others and, thus, he as well as others may be using the term 'pain' to refer to completely different things – or to none at all. Indeed, he invites us to assume that in the language he shares with others it would be perfectly sensible so to say and suggest. However, it does not make sense to suggest that we may talk about, let alone *determine* any individual differences in our feelings of pain or in our use of the term 'pain', *unless* the implication of the term 'pain' is shared by everyone involved, and unless, furthermore, this term is used to refer to the same sort of "thing". Without these presuppositions, any discussion of the question of 'pain' disintegrates into nonsense.

In summary, it would seem that the question of 'pain' as set out by Wittgenstein is *obtuse* in the sense that putting it forward presupposes that we know the meaning and use of the terms of the question and also to what it may be correctly applied. But then we are asked to forget or ignore this presupposed knowledge, and to pretend that it is immaterial for a discussion of what the terms 'pain' may be correctly used to refer to – or whether it may be used correctly to refer to anything at all. But it is a question which can only be asked and discussed granted we have already learned a language which we may use to talk about pain, and thereby granted pain to be the kind of "thing" which exists as something we may talk *correctly* about and *refer* to. Hence, if we do ask this question, the answer is logically implied: an affirmation would be redundant, while a denial would be contradictory.

Now, if the same sceptical question had concerned the use of the terms 'cups' or 'neurons' instead of 'pains', it would have been obvious why such a question would not make sense. Indeed, it would probably have been so obvious that we would hesitate to ask it. For if we did ask this question about the use of the terms 'cups' and 'neurons', we might as well ask the same question about all other terms in our language – and we would be well on the road to asking whether we can be sure that we may use terms in our language to refer to and talk correctly about *any* objects in material reality. However, although in *particular* cases we may be in doubt as to whether a particular term may be correctly used to refer to some particular thing, i.e. a thing which has been identified in a shared public world, we cannot doubt that as language users taking part in this discussion, we do know (other) correct terms for the thing (i.e. those forming part of the identification of the thing), nor doubt that we know how to use *these* terms correctly. One cannot doubt the necessity of these conditions for settling the question under discussion – unless, of course, one has been seriously contaminated with scepticism, and mistakenly assumes that one may get away with using language to question the very possibility of using language to talk correctly about anything.

However, similar conditions seem to apply to the question of 'pain' – and by extension to questions concerning other internal or mental states such as emotions, thoughts and belief – in the sense that scepticism as to whether we may use language to talk correctly – or at all – about such states, and make reference to them, presupposes that we know the meaning of terms for these states, and that together with other language users we may determine what they may be correctly used to refer to. So rather than questioning the existence of such internal or mental states and the possibility of having knowledge about them and of correctly describing them and being able to communicate about our experiences of such states, this questioning logically rests on the assumption of both their existence and this possibility. Indeed, without these assumptions concerning the *intersubjectivity* of both our experiences of, and of our use of terms to talk about and refer to our non-publicly observable internal or mental states, neither ordinary everyday communication nor philosophical discussions about such states would be possible.[2]

In the section which follows I shall further clarify what it means that our cognition and use of language is *intersubjective*, just as I shall show that the very same assumption of intersubjectivity is a necessary condition for knowledge, descriptions and communication of both that which is and that which is not publicly observable and shared.

The Intersubjectivity of Public and Personal Knowledge and Experiences

It has to be admitted that it is somehow puzzling that pains and other internal states, which are only directly observable to the persons who have them, and are not *observationally* shareable by others, are nevertheless things which we may communicate about in a language we *do* share with other persons. So, apparently, shareability in the sense of being *publicly* observable and known cannot be a condition for the possibility of communication among persons about things which only they may directly observe, experience and know about.

In the case of observing, experiencing and communicating about objects in material reality, our situation is arguably significantly different. Take for example two people sitting on either side of a table with cups and plates, a teapot, a bowl of sugar and a bottle of milk. All these things exist in a shared public world and are perfectly observable to both persons; they may together determine the things

---

[2]I am not saying that the fact that we have a language with terms for both objects in reality and pains in bodies *proves* the existence of objects in reality and pains in bodies, nor that all and every concrete statement we put forward about either objects or pains are always or infallibly correct. What I am saying is that we cannot begin to discuss or investigate language and the use of language to talk about such things as objects or pains without assuming, generally, that both objects in reality and pains in bodies exist as things that we may have knowledge of and use language to talk correctly about. One cannot take part in this kind of philosophers' discussion without committing oneself epistemologically.

on the table and also whether the descriptions they put forward about them are correct. And yet, it could be argued that how these things are observed and appear to them from their different vantage points on either side of the table is different, i.e. due to the fact that the parts and features of the things which are directly observable to the one, are not the same as those which are so observable to the other. However, this does not present any serious difficulties since, first, it is part of our knowledge of things having been identified as particular things, that they will appear differently when e.g. we move around and look at them from different vantage points – and that, generally, things do appear differently when observed with regard to different possibilities of observations and action. And these differences of perception and experience do not represent any serious problems since, secondly, none of the particular ways of perceiving and experiencing the things on the table, and none of the descriptions by either person of their experiences of the cups, plates, etc., are unique to him or her. Indeed, it is *assumed* that they are not, just as in general any person and language user will assume that if other people could look at the things from his or her vantage point(s), they would observe what he or she does, and report that they perceive the same features and properties of the things, and describe what they perceive as he or she does. *If we could not count on this assumption, communication and action between persons about things in the world would be impossible*.

But there are numerous other ways in which the knowledge and description of persons concerning things in publicly observable reality may differ. Just think of the differences due to our different background, education, previous history of experiences, and the opportunities to observe and describe such things which are or have been available to us. Examples are *legion* – I only have to think of the knowledge I have of aeroplanes, their construction and how to fly them compared to that of a pilot. Or, conversely, think of the knowledge I have as a clinical psychologist about the transference phenomena occurring between client and therapist during psychotherapy compared to that of an aeroplane pilot, who has not encountered such phenomena, and who does not have the language and terms to describe them that I have. And yet, despite the fact that our knowledge of these and numerous other matters are not exactly the same, and probably never will be, we are in fact able to make available and to communicate the knowledge that each of us has about aeroplanes and transference problems and those other matters, and thus to share our knowledge of what in this respect is personal to each of us. But if the condition for the intersubjectivity and shareability of knowledge and description in the actual case as well as between persons in general, is not and cannot be that persons have exactly the same knowledge and experience of things, or the same possibilities of describing things, since this condition is only rarely if ever met due to their different background, education, history of experiences, and so forth, on *what* then relies this intersubjectivity of knowledge and description of persons?

It relies no doubt on the fact that, apart from differences in our knowledge, experience and background, we do share a substantial amount of knowledge and descriptions of the world in which we live and act, of the things with which we may act, of ourselves and of the persons with whom we may co-act. However, to say so does

not of course add anything to our notions of 'shared knowledge and description', that is, knowledge and descriptions of things which have been available to shared inspection. Nor does it suffice to account for how knowledge of things which may only have been available to ourselves, and which in this sense is personal, is amenable to description and intersubjective communication. It does not do so unless it is assumed that the vantage points, backgrounds, and situations we may be in, are *in principle* shareable, and thus that other persons *could be* or *could have been* in the same situations. Nor does it suffice independently of assuming that, granted other people had had the same background, or had been in the same situations that we ourselves have been or are in, then they would have the knowledge of the situation and the things that we have, and describe them the way we do. However, this suffices, indeed it will *have* to suffice to say that it is fundamental to the cognition and experience of persons that, although other people may not be in our situations, and may not have, or may not have had, exactly the same experiences and knowledge that we have or have had, they would – *could* they be, or *had* they been, in our situations. Likewise, it suffices, and will *have* to suffice to say that to be language users and to share a language with other persons *logically* implies and presupposes that other language users, granted that they could be in our situation and have the experience, knowledge, background, points of view etc. that we have, would use language to describe what we experience, know of, etc., in those situations as we do. Or, they would consent that the descriptions we put forward about our experience and knowledge are correct and correctly applied.[3]

Now, if we can agree that these presuppositions must be fundamental to the cognition, use of language and communication of persons, and indispensable for any meaningful discussion among persons about what they know and how they describe what they know, I think we shall also have to agree that this intersubjectivity of cognition and language relies on a notion of 'truth' which implies that what is true or false, correct or incorrect, is also true or false, correct or incorrect for other persons. That is, this intersubjectivity relies on a notion of 'truth' which *logically* implies and presupposes a notion of 'others'. (I shall expand on this point in a later section.) This assumed intersubjectivity must be the rock bottom, the very point of departure from which any discussion about our knowledge and description must be based and proceed – whether such discussions concern our knowledge and description of thing in publicly observable reality, or our internal states, such as our emotions, thoughts or feelings of pain.[4]

---

[3]The importance of this intersubjectivity of human cognition, communication and co-action, becomes clear when we consider that a substantial part of our knowledge of both material reality and of the societies in which we live and co-act with others, does not rely on first hand personal experiences, but rather is knowledge we have adopted or acquired from others. In this sense knowledge of reality thus acquired resembles knowledge communicated to us by others about their mental states, and in the sense, furthermore, that both kinds of knowledge relies on and presupposes the conditions of intersubjectivity of cognition and use of language outlined above.

[4]As argued extensively elsewhere (Praetorius 2000), this assumed intersubjectivity of cognition, language and of the notion 'truth' can neither be proved nor doubted without being conceded, and hence will have to be taken granted as a *principle*.

In what follows it will be argued that the intersubjectivity of human cognition, language and of the notion of 'truth' is a condition for the possibility of rigorously determining and thus of distinguishing between knowledge and descriptions of that which is publicly observable and that which is not – and hence for people together to develop conventions, criteria or standards for the validity and objectivity of their knowledge and description of things and events in publicly observable reality. And it will be argued, furthermore, that on these conditions relies the possibility of characterizing and determining individual differences and variations in the cognition and description of different persons – both concerning that which is publicly observable and shared and that which is *personal*.

Conditions for Distinguishing What Is Publicly Observable from What Is Not

It is well know that considerable variations may exist in how things are described by different persons, and by the same person in different situations, not only when the things concerned are "internal" states and events, but even things in material reality. Not only our intentions, purposes and possibilities of observation and action, but even our moods and temperaments may at times determine what we experience, and how we experience and describe ourselves and the rest of reality. So, would it not then be reasonable to reserve the notion of 'true descriptions and assertions', and 'knowledge of what exists objectively' for cases in which no such individual differences and variations exist, and to reserve the terms 'exist' and 'determinable' for things and events about which no uncertainty prevails – because they belong to what is materially and publicly observable?

This solution has been attempted, notably by the logical positivists and by the radical and logical behaviourists in both psychology and philosophy, who aimed to establish a firm epistemological basis for scientific research. In this pursuit they argued that only *that* exists objectively, and hence can be the object of scientific research, which may be determined by rigorous public criteria and standards, and that only knowledge and description of what has been thus determined and observed, may be said to be meaningful and true. In effect, any determinate notions of the truth and meaning of statements and knowledge would have to derive from observations fulfilling such criteria and determinations. Consequently, what cannot be subject to rigorous public scrutiny and observation fulfilling such standards and criteria does not exist, nor can description of things which cannot be so observed be true; hence, descriptions and the existence of knowledge of such things may be discarded as nonsensical.

However, against such arguments we only have to consider that the very process by which we identify and determine a situation, and what in this situation is materially and publicly observable, presupposes that something is the case or true about the situation and things being determined and observed, which is also the case or true to other people – *in casu* the people involved in the observation. Thus, it is not because

situations exist or may be arranged, in which things are publicly observable, and which we may come to agree to describe in particular ways, that the notion "emerges" of what – for everyone involved – is true or correct about things and situations. It is the other way around – for no such determinations of correctness of knowledge and description of things and situations could be agreed upon, let alone be arranged and function as criteria or standards for correctness, unless it was presupposed that *when* arranging and determining these criteria or standards, we *already* have a concept of 'truth' which we know how to use correctly; a concept of 'truth', furthermore, which is such that what is true or false, correct or incorrect, is also true or false, correct or incorrect for others.

Thus, the point, so easily overlooked, is that even in a situation in which the things and events being described are publicly observable and identifiable, it is logically implied and presupposed of descriptions put forward and being understood by others, and of these others, being in the same situation in which we are, that they will have the knowledge we have about these things and events, and will describe them as we do. In general, it is presupposed that our notion of 'correct' or 'true' is such that what is true or correct, is also true or correct for others. *This* presupposition granted, it is possible for persons – in spite of the differences and variations which exist in their knowledge, background, opportunities for observation and action – to arrange conditions and develop criteria and standards for what may count as correct, *objective* descriptions and *publicly* existing things, that is, possible *to determine and distinguish between conditions under which such criteria or standards apply, and in which they do not*. Furthermore, it is *because* of this presupposed intersubjectivity that it is possible to determine individual differences in the cognition and descriptions of different persons – both concerning that which is publicly observable and that which is only observable to the persons themselves, such as their mental or internal states. In other word, it is because *both* the knowledge shared by different persons *and* the knowledge of persons which is personal, are perfectly sensible issues of intersubjective discourse *that it is possible to distinguish between what is publicly observable and shared and what is personal*.

Let me illustrate this point by giving the following example. One of my friends tells me: "I am terribly depressed; everything looks so grey and colourless – even the trees and flowers look grey and colourless". Now, could it not reasonably be argued that at least in this case we are not talking about a "public" issue, but rather of something "private", and also that it is a situation in which it would make no sense to maintain that my friend is still using language correctly? Is it not a situation in which any well-defined notions of correct or true assertions has been suspended? Not at all. For one thing, I do understand what my friend is saying. I am perfectly able to communicate with him about his – in this case – curious experience of the colours of trees and flowers. But a condition for maintaining that I understand what he is saying, and for communicating with him about his curious experiences of the colours of trees and flowers is, naturally, that he still uses language correctly when talking about his experiences of these things. That is, it is a condition that he knows the correct implications of terms for various colour categories, and that he knows how to apply them correctly. And it is a condition that what he is talking about is something

he may refer to, and about which true and false assertions may be made, i.e. his curious experience of the colours of trees and flowers. Thus, it is a condition that both of us are still using the same language – indeed the very same language that he and I use under normal circumstances to talk about quite ordinary everyday matters; and it is a condition that we are together able to determine *what* he is talking about. If we are able so to do, and thus able to talk about his experiences, however curious, it has to be maintained that he is using language correctly when describing his experiences.

However, it is quite clear that his description of the colours of the trees and flowers is not of *general validity*, and I do not take his description as an attempt on his part to produce descriptions of general validity – i.e. descriptions which would be correct under normal everyday conditions of observation. But an important part of the message he is trying to convey to me – and which I understand – is precisely that his situation is *not* normal, i.e. that his depression affects his perception and description of things in ways which differ from how he normally perceives and describes them. A fact of which he himself is perfectly – and probably painfully – aware.

So, although my friend may feel eminently on his own with his unusual and personal experiences during his depression, neither his experiences nor his descriptions of them are *private*. They are perfectly understandable to others because it is presupposed – by us and by him – that he is using language when describing what he experiences as others would – could they be in his situation and experience what he does. Indeed, our communication about what he experiences relies on the presupposition that what to him is the case or true about his experiences would also be the case or true for others, had they his experience.

Let me conclude my discussion of this example by saying that it shows that individual differences in the cognition and description of reality of different persons may indeed exist and be determinable. However, it also shows that a condition for these differences between persons and their cognition and description to exist, is that persons and language users, despite such differences, share a vast amount of knowledge and correct descriptions of reality. And it shows, furthermore, that the possibility of determining and of talking correctly about such differences relies on the presupposition that, when we describe what we know of or experience, we use language to describe it as others would, if they had our knowledge and experience; however, this in its turn both presupposes and implies that the notions of 'correct' and 'true' of persons are such that what is correct or true, is also correct or true to others. If these presuppositions and implications concerning the intersubjectivity of our cognition and language did not apply, no personal differences in knowledge and experiences could exist, nor any possibility to determine or talk sensibly about such differences – whether these differences concern what is or what is not publicly observable. Conversely, granted these presuppositions, and hence that situations, points of views and observations are in *principle* shareable, it is possible for persons to determine and distinguish between individual differences and variations in their cognition and description concerning both what is and what is not publicly observable, and to determine under what circumstances such individual differences occur. And it is possible *therefore* for persons together with others to develop

criteria and standards with which rigorously to determine and distinguish between first-person and third-person observation and description, so crucial for the possibility of developing sciences.

In the section which follows I shall clarify what it means that the intersubjectivity of cognition and language relies on a notion of "truth" which logically implies a notion of "others".

Implications of the Logical Relation Between the Notion of "Truth" and "Others": the Impossibility of Private Cognition and Languages

Suppose my friend's condition has deteriorated; he now tells me: "I have experiences and knowledge about some particular "things", but what I take to be true or correct about them, is not true or correct for others, – or I cannot be certain that it is". Now, what could he possibly mean by that? Could what he says mean, for example, that if others could experience and know of the "things" that he experiences and knows of, then what for them would be the case or true about them, is not the same as what for him is true or the case about them? Well, if that is what he means, he is obviously contradicting himself, for in that case it would be impossible to ascertain that "the things" being experienced and known by others concern the same "things" as those experienced and known by my friend. Or, could it be that he means that these "things" of which he knows and experiences, are in principle inconceivable to others, because the notion of true and correct in the case of his unique experiences and knowledge of these "things" is different from the notion of true and correct which others have, and which he himself has in other cases, for example when communicating to others his unique experiences and knowledge. In other words, does he mean that this concept of true or correct, which applies to his experiences and knowledge about these particular "things", is a concept which is special to him, in the sense: *private*, and consequently, that his experiences and knowledge of these "thing" are equally private?

Now, for such a claim to have any bearing – even for himself – it would seem to require that he be able to account for how his "private" concept of true and correct differs from the one he shares with others, i.e. account for it in the language in which the claim is put forward. But if he could do that, his "private" concept of true and correct would *not* be private, nor inconceivable to others. Likewise, in order to maintain that what he experiences and knows to be the case or true about these particular "things" is not the case or true for others, would seem to require that he be able to determine what would be the case or true for others about these "things", and how it differs from what is the case or true to him – again in the language in which the claim is made. However, if he could do that, then what to him is true or false about the "things" he experiences would be perfectly conceivable for others, and what he knows about them would be perfectly expressible in terms of the language which he shares with others.

So, we may conclude that my friend is either contradicting himself or talking nonsense – or both. This would have been immediately obvious if instead he had said "I have discovered that the statement, "…", is true or correct, but it is not true or correct for others, or I cannot be certain that it would be true or correct for others". What is obvious is that he mistakenly thinks that one could share a language with other people, and also that in this language the notions of 'true' and 'correct' could be different for different people.[5]

I think we shall have to agree that for the same reason that nobody may claim to possess private knowledge or a private language, and thus a private notion of truth, no such private language or knowledge may be ascribed to others – neither *in toto*, nor in part. For, how could we justifiably ascribe a private language or knowledge to others without being able to account for both such a language and knowledge in a language which is *not* private, and into which this knowledge and language must somehow be translatable? This, I think, suffices to show that *one cannot know for oneself what others might not in principle know as well*; that is to show, once again, that to be a user of a language one shares with others, means that one cannot know what is true or correct to say about a "thing", which may not be true or correct to others, could they experience and know what one knows about the "thing".

What It Means That Our Notion of "Truth" Is Fundamentally Social

The previous section argued that the intersubjectivity of cognition and use of language of persons to describe what they know about themselves, others and the world around them relies on a notion of "truth" which logically implies a notion of "others", and hence on a notion of "truth" which is fundamentally social. Now, it would seem to be almost self-evident that to be able to communicate about what one knows, implies knowing that one's knowledge, categories, conceptual systems and descriptions are indeed "inter-personal", i.e. that one shares such categories, conceptual systems and descriptions with others. Although I may know of and describe things which others do not (yet) know of, or know different things about them than others do, to know and to say so necessarily implies that if others had the possibilities of observing and describing the things that I have, then others would know what I know about them and describe them the way I do. In other word, what to me is true or false would also be true or false to others.

However, the notion that the 'truth' of cognition and language of persons is fundamentally social does *not* mean that it relies on the fact that persons may come to agree on and make conventions, and develop criteria, standards or rules for what may "count" as objective and true knowledge and descriptions of the things or situations in the world in which they find themselves. On the contrary, no social

---

[5] There is of course the possibility that what he means is merely that he is not sure of the correct implications and application of the statement – but that is a quite different matter.

conventions, agreement or criteria about the truth and objectivity of knowledge and correctness of descriptions could be established among persons, unless *prior* to establishing such conventions, agreement and criteria, they had together determined and identified *both* things and situations to which these conventions and criteria apply, and *therefore*, unless they already had a notion of 'truth' in which it is presupposed that what is true or correct is also true and correct for others. Hence, to say that our notion of 'truth' is fundamentally social does not mean that the notion of 'truth' is a social phenomenon, i.e. a product of socially agreed practice. On the contrary, social phenomena and practice, including the development of conventions, criteria, rules or agreement on how to use language and its terms, depend on notions of 'correct' or 'true' and 'incorrect' or 'false' which are inherently shared.

Consequences for Developmental Psychology: Conclusion

However fundamental – and almost embarrassingly banal – the presuppositions concerning the intersubjectivity of cognition, language and the notion of truth of persons may seem, it has been widely overlooked within philosophy of mind and consciousness. According to the traditional assumption, shared by many philosophers even today, we all start out as "Cartesian subjects", having knowledge and experience of the content of our own mind, i.e. our sense data, perceptions, thoughts, emotions and feelings. From this supposed private, though certain knowledge and experiences "from our own cases" it is believed to be possible to work towards true knowledge of the nature of what causes this content and the rest of the objective order of reality, including other persons, and to develop a language in which we may talk with others of this knowledge and experience. This same assumption also seems to inspire and lie behind functional models and accounts of the cognition and use of language of people currently being developed within Cognitive Science – be those models computational *cum* representational or connectionist. However, it would seem to be a insoluble problem for such models to account for how the cognition of an individual – formed in "splendid solipsistic isolation", and with no notion of the intersubjectivity of its own cognition and that of others – would ever come to accord with the cognition of others. And it would be impossible without such a notion to explain, moreover, how they come to distinguish that part of their knowledge concerning themselves and the world which is *personal*, from that which is *public* available and shared by others.

George Herbert Mead was among the first within psychology to provide substantial arguments against the assumption that individuals living mentally in isolation from others, could ever develop or acquire knowledge about themselves, others and reality of the kind and in the way that we humans actually do, and in particular to argue that no such individual could be aware of itself – since, according to Mead, to be aware of oneself is to "look" at oneself from the standpoint of another (Mead, 1934). In view of the arguments presented so far, it seems that we are now able to strengthen Mead's original claim. If we suppose with Mead that to

be aware of oneself requires being able to "look" at oneself from the standpoint of another, it has to be conceded, that this "looking" both presupposes and implies that if one could be in the position of other persons and look at oneself as they do, then one would see and come to know oneself the way others do. Without the presupposition that what, from the position of others is known by them to be the case or true about me, would also be the case or true for myself, if I could be in their situation and have their knowledge, there would be no sense in talking about seeing oneself from the point of view of others. Thus, a notion of truth which is such that what is true or the case is also the case for others, and hence a notion of truth which logically implies a notion of others, must necessarily *precede* the possibility of anyone "looking" at himself from the standpoint of another; it is not something which may be acquired by such "looking", nor by imagining such "looking".

Apart from these logical reasons, there seems to be empirical support from developmental psychology for assuming that for an individual, say a child, to become aware of himself and of others as persons, and of acquiring notions of his or her own mind and cognition and those of others, it is necessary that he or she be *received* and *understood* by others as a person. Thus, empirical research of early mother–infant communication seems to show that a child's ability to develop knowledge about himself, reality and other people around him, and later on to acquire a language to communicate this knowledge and that of others, depends on the mother's (or other care-person's) indefatigable effort and willingness to understand and see the child's behaviour as being *intentional*. And it seems to rely on the mother's efforts to understand, not only what goes on "inside" the child, but to interpret the child's reactions to her, and his action with things, as expressing his knowledge about things, and his attempts intentionally to act upon them (for an excellent account of the development of early mother-infant interaction and communication, see Bruner, 1983).

Empirical investigations also seem to show that the child up to a certain age – presumably due to an over-generalisation of what he sees as his mother's apparent unlimited knowledge about his experiences, intentions, needs and actions – believes that others are in the same situation as he himself and share his point of view, and the knowledge he has about things in those situations – and even that they may "have" his thoughts and feelings. Thus, the child seems to over-generalise the fact that "what he knows may also be known by others" to mean that others do indeed find themselves in exactly the same situation as he does, and having the same knowledge he has, and having access to his thoughts and feelings. Only later on does the child learn that other people may perceive the same situation from points of view which are different from his. And only later does the child realise that part of himself, his feelings and thoughts, are only directly observable to himself, and also that this part of him is what makes him uniquely *him*, being a person both physically and mentally distinct from other persons.

If we go back to the assumption held by most philosophers even today, namely that the child starts with "private" and "subjective" knowledge about himself and the world, it would now seems that that this assumption turns the issue on its head. For, if what empirical investigations of the child's initial development of knowledge

seems to indicate is correct, and what from a logical point of view must necessarily be the case, the child does not start with private knowledge "from his own case", but with knowledge of which it is assumed by the child that it is indeed shared by others – i.e. by his mother. The problem, it would seem, is rather to account for how the child later in his development comes to appreciate that, although what he knows may indeed be known and shared by others, others may not be in exactly the same situation as he himself. Thus, the problem seems to be to account for how the child begins to learn to appreciate the notion of 'different points of view', and how he begins to learn and appreciate the difference between what is and what is not observable to him as opposed to others, i.e. that the knowledge one may have of a situation and of oneself may be *personal*.[6]

However, what cannot be accounted for nor explained, but which has to be presupposed and taken for granted, is that for a child to be able to learn this from other people in the community in which he grows up, and be able to take part in their "forms of life", the child must have a notion of 'truth' which is such that what is the case, true or false, is also the case, true or false for others. What is lacking in epistemological approaches which start from the position of the individual alone set against the rest of the world – be they generic constructive approaches, or biological or computational functional approaches – is not just a social context of others, which enables the individual to confront and compare his knowledge with the knowledge of others for the purpose of determining, for example, whether his knowledge is in accordance with theirs, and thus may "count" as objective or true, or whether it relies on one's subjective dreams, illusion or imagination. What is lacking is precisely a notion of 'truth' which logically implies and presupposes a notion of 'others', which makes it possible for persons together to develop procedures for determining the objectivity and truth of their knowledge, and for everyone to compare his knowledge with the knowledge of others. That is, an intersubjective notion of 'truth' which makes it possible for someone to *be* a person, i.e. someone who may share an incredible amount of knowledge with other people, but who also has knowledge about himself and the world which is uniquely personal – and thus is someone who also differs from other persons; someone with whom we may agree – and disagree.

References

Bruner J (1983) Child's talk. Learning to use language. Norton, London
Mead GH (1934) Mind, self, and society. University of Chicago Press, Chicago, IL
Praetorius N (2000) Principles of cognition. Language and action. Essays on the foundations of a science of psychology. Kluwer, Dordrecht/Boston/London
Tomasello M, Rakoczy H (2003) What makes human cognition unique? From individual to shared to collective intentionality. Mind Lang 18:121–147
Wittgenstein L (1945/1953) Anscombe GEM, Rees R (eds) Philosophical investigations. Blackwell, Oxford

---

[6] For thorough investigations of this development, see Tomasello and Rakoczy (2003).

The Problem of Representation

Michael Wheeler

How Things Look

As we shall be thinking of it here, to engage in representational explanation is to explain behaviour by invoking internal states with content (i.e. internal states that encode meaning or bear information). Thus one may account for why Voldemort triumphantly parades Harry's limp body in front of the assembled fighters at Hogwarts by explaining that Voldemort possesses a behaviour-influencing inner state that bears the (as it happens) inaccurate content, 'Harry is dead'. Here, then, is how things look to the representationalist. Intelligent agents take the world to be a certain way. One animal may represent a second animal as a potential dinner, while the second may represent the first as a threat. Of course, as Voldemort discovered to his detriment, the world need not be the way agents take it to be. Philosophers often develop this thought by saying that where we encounter an agent that is genuinely capable of representing the world, there already exists the possibility of that agent *mis*representing the world.[1]

What exactly does cognitive science add to this picture of agent-world relations, a picture that is deeply embedded in commonsense (at least in modern times) and has a venerable philosophical ancestry in the work of Plato, Descartes, Hume and Kant, among others? The answer is a direct route to the naturalization of representation, via a seemingly compelling story about what it is at the level of bodily machinery that makes it possible for representing agents to take the world to be a certain (possibly

M. Wheeler
Department of Philosophy University of Stirling

[1]Some recent treatments of representation (e.g. Rowlands 2006; Gallagher 2008) illuminate the concept by way of more detailed lists of properties that representations are standardly taken to have. Candidates for inclusion on such a list might include passivity, having duration, requiring interpretation, and decoupleability from the object or state of affairs that is being represented. In the case of some of these properties (e.g. having duration) I believe they are entailed by the more open characterization that I have given. In the case of others (e.g. decoupleability, see note 6 below) I believe they are not strictly required by the basic idea of representation, even if they are often present. Anyway, for the purposes of the present investigation it will be strategically advantageous to capture the essence of the notion in the most general way, which is what I have tried to do.

incorrect) way. According to that story, inside the agent's head – physically inside the agent's skull, that is – there exist systematically organized (ultimately) neural states whose functional role is to *stand in for* (usually external) objects and situations in the agent's behaviour-producing neural mechanisms. At root, it is on the basis of this standing-in-for relation that the states in question acquire the status of internal vehicles of content – that is, of inner representations.

So construed, representations have always been rather dear to cognitive science. Indeed, it is arguable that the notion of representation is one of the conceptual keystones on which the field has been built. Thus mainstream cognitive scientists, classicists and connectionists alike, hold not merely that the concept of representation often makes an important contribution to good scientific explanations of many psychological phenomena, but that explanatory strategies that appeal to representations offer our *only* hope for a scientific understanding of psychologically interesting behaviour. Of course, classicists and connectionists famously disagree about the form that representations take. Classicists, inspired by natural and artificial languages, favour symbolic structures with a combinatorial syntax and semantics (see, e.g. Newell and Simon 1976). Connectionists, inspired by the abstract architecture of the brain, favour shifting patterns of activation distributed across interconnected groups of simple processing units (see, e.g. Hinton et al. 1986). Fur has been known to fly (e.g. Fodor and Pylyshyn 1988; Smolensky 1988). But however tetchy things get, neither orthodox camp claims to dispute the presence of representational structures nor the power of representational explanation. One might think that things must have changed recently, with the advent of so-called embodied-embedded cognitive science and its stress on the subtle ways in which intelligent action emerges from real-time dynamical couplings between brain, body and environment. And it is true that this development in the field has sometimes provided a platform for scepticism about representation (e.g. Varela et al. 1991; Thelen and Smith 1993; Webb 1994; Cliff and Noble 1997; Gallagher 2008). Nevertheless, the fact is that most (although admittedly not all) embodied-embedded theorists have ultimately remained remarkably unpersuaded by the voices of anti-representational dissent. As cognitive scientists we continue to live, it seems, in an impressively stable age of representation.

It is worth driving this point home by pausing for a moment to consider Robert Cummins' distinction between two different issues that arise in the vicinity of representation. Cummins (1996, p. 66) observes that one problem the representationalist faces is how to 'pin the meaning on the symbol'. This problem has been much discussed, but it is not strictly fundamental. For it assumes that representational language is essentially in order and that 'all' that is required is an account of how to specify the content of the already-known-to-be-representational-in-character states. The logically prior problem (identified by Cummins) remains: why and when should we appeal to representations (content-bearing inner states or vehicles) *at all* (see also Vacariu et al. 2001)? Call this *the question of representational warrant*. It is a question that, across cognitive science as a whole, is hardly ever asked, or at least not in a way that invites more than a few cursory observations about (a) the fact that intelligent action is typically sensitive to the information carried by

environmental stimuli and not merely to the physical form of such stimuli, and/or (b) the aforementioned possibility of epistemic error.

Given all this it is striking that the really interesting engagement between phenomenology and cognitive science over the issue of representation places the question of representational warrant centre-stage. The task of this piece, then, is to show how one well-established stream in phenomenological thought about intelligent action (a stream that starts with Heidegger and flows via Merleau-Ponty into recent writers such as Dreyfus) provides an answer to the question of representational warrant that ultimately casts doubt on the pivotal psychological role that cognitive scientists of all stripes usually reserve for the concept of representation.

Encountering Entities

In Division 1 of *Being and Time*, Heidegger (1927) argues, on phenomenological grounds, that we ordinarily encounter entities as (what he calls) *equipment*, that is, as being *for* certain sorts of tasks (cooking, text-editing, baby bouncing, and so on). He also suggests that we achieve our most fundamental relationship with equipment not by some detached intellectual or theoretical study of the entity concerned, what we might think of as an *offline* engagement with it, but rather by skilfully manipulating it in a hitch-free manner, what we might think of as an *online* engagement with it.[2] Entities encountered in action in this way have their own distinctive mode of intelligibility that Heidegger calls *readiness-to-hand*. This is marked out by a distinctive phenomenological signature. While engaged in hitch-free skilled activity, the manipulating agent has no conscious experience of the item of equipment in question *as an object* (as a bearer of determinate properties that exist independently of the agent-centred context of action in which the entity is involved). Neither does that agent have any experience of herself *as a subject* (as a self-sufficient entity existing over and against her context of action and the equipment with which she is involved). For example, while engaged in trouble-free text editing, the skilled typist has no conscious recognition of the keyboard or the monitor, in the way that one would if one simply stood back and thought about them. Moreover, the skilled typist becomes absorbed in his activity in such a way that he has no awareness of his existence as a self over and against his ongoing activity. Importantly, it does not follow from this that the agent's behaviour in such smooth coping scenarios is automatic, in the sense of there being no awareness present at all, but rather that the kind of awareness that is present (what Heidegger calls *circumspection*) is non-subject–object in form.

Although Heidegger rejects the idea that the categories of subject and object characterize our most fundamental way of encountering equipmental entities, he claims that the subject–object dichotomy does characterize a derivative kind of

---

[2]For more on the online-offline distinction, see Wheeler and Clark (1999), Wheeler (2005a, b).

encounter. When the agent engages in, for example, theorizing in natural science, or when sensing takes place purely in the service of reflective or philosophical contemplation, the entities under study are phenomenologically removed from the action-oriented settings of everyday equipmental practice and are thereby revealed as fully-fledged objects, that is, as the bearers of certain context-free determinate or measurable properties (size in metres, weight in kilos etc.). Heidegger calls this mode of intelligibility *presence-at-hand*. The emergence of entities as present-at-hand is accompanied by the emergence of the agent as a subject over and against an objective world. Encounters with the present-at-hand are thus fundamentally subject–object in structure.

Sandwiched phenomenologically between readiness-to-hand and presence-at-hand is a rather less explored (by most commentators) mode of encounter that Heidegger calls *un-readiness-to-hand*. It emerges when skilled practical activity is disturbed by broken or malfunctioning equipment, discovered-to-be-missing equipment, or in-the-way equipment. When encountered as un-ready-to-hand, entities are no longer phenomenologically transparent. However, they are not yet the fully fledged objects of the present-at-hand, since their broken, malfunctioning, missing or obstructive status is defined relative to a particular equipmental context. For instance, if I am on the way to work a broken watch is not encountered by me as a lump of metal of a measurable mass, but rather as a faulty timekeeper, a damaged item of equipment that constitutes a hindrance to my train-catching activity. Correlatively, disturbances reveal the agent in the mode of a practical problem solver. In most cases that problem solving will remain an online achievement involving embodied interactions between agent and world (e.g. the switching of activity to use different time-related resources in the environment). This is a domain in which intelligent agents produce richly adaptive, fluid and flexible, real-time, context-sensitive embodied responses to environmental problems. In the limit (e.g. when a watchmaker uses her theoretical knowledge of how watches work to guide a repair), the agent's problem solving activity will approximate the theoretical reasoning distinctive of science. But even then the agent is not 'just theorizing' or 'just looking', so it is not yet, in Heidegger's terms, a *pure* disengaged subject. In sum, as smooth coping is disrupted and we enter the realm of un-readiness-to-hand, a cognitive distance is introduced between agent and entity, a progressively increasing distance (with progressively disruptive disturbance) that marks a phenomenological journey from a minimal to a full-blown subject–object dichotomy.

What has all this got to do with representation? Here we need to go beyond Heidegger's own analysis. On the one hand, it is hard to see how to make sense of an agent enjoying psychological *re-presentations* of its world unless that agent is already in some way understood to be a subject over and against an independent world of objects, with the cognitive distance between agent and world that such an arrangement implies. On the other, and more importantly for the present argument, where there exists a subject–object dichotomy, there exists the issue of how the subject gains epistemic access to the independent world of objects that it inhabits. Although this issue does not perhaps necessitate a response couched in representational

terms, it certainly invites one.[3] If we couple these points to Heidegger's tripartite phenomenology of encounters with entities, we get the following result which amounts to an answer to the question of representational warrant. Smooth coping (hitch-free skilled action, readiness-to-hand) is not characterized by any sort of subject–object dichotomy, whereas both disengaged theorizing (presence-at-hand) and practical problem solving (un-readiness-to-hand) are. So smooth coping is non-representational in form, whereas disengaged theorizing and practical problem solving warrant representational explanation.

Crucially, Heidegger's phenomenological analysis suggests that disengaged theorizing and practical problem solving (especially at the fully online end) involve different kinds of representational states. When revealed as present-at-hand, an entity will be experienced in terms of properties that are action-neutral, specifiable without essential reference to the representing agent, and context-independent. On the strength of the co-dependence that Heidegger finds between agents and entities in the manner in which they are disclosed, it seems that this group of entity-related properties will also characterize the contents of the theorizing subject's representational states. So, for example, the disengaged reasoner, reflecting offline on a navigation problem, may represent the external environment by way of something like a Cartesian co-ordinate system in which the objective shape, orientation and positions of detected obstacles are plotted. By contrast, when revealed as un-ready-to-hand, an entity will be experienced in terms of properties that are action-specific, egocentric and dependent on a particular context of activity. This second group of properties will also characterize the contents of the practical problem solver's representational states. So, for example, the online, task-engaged navigator may represent the external environment by way of an egocentrically defined space in which obstacles appear only as regions to be avoided, positioned in terms of roughly specified bearings relative to her own body.

This concludes our first foray into Heideggerian phenomenology. We'll be revisiting it later. First we should say what Heidegger's tripartite analysis of our encounters with entities tells us about representational explanation in cognitive science. For that, some brief stage-setting is required.

Agents and Their Parts

There's a well-worn worry that cognitive psychologists, schooled in the unreliability of introspection, have about phenomenology, namely that it's at best a dubious guide to the mechanics of our cognitive processing. There are good reasons to

---

[3]Notice that the problem that cries out for a representational solution here is not that of how an individual is able to distinguish itself from its world, but that of how an individual is able to gain competent and appropriate epistemic access to its world, given that it is distinguished from that world precisely as a subject distinguished from a collection of independent objects. Thanks to Shaun Gallagher and Matthew Ratcliffe for helping me to get clear about this point.

resist this view, but it needs to be rejected with care, not impatience. Of course, phenomenology in the Heideggerian tradition is not equivalent to introspection. Rather it's a theoretical enterprise that through an attentive and sensitive examination of ordinary experience, aims to reveal the transcendental yet historical conditions that shape and structure that experience. Still, that adjustment is unlikely to satisfy the scientifically minded sceptics, so what might? Here it is useful to look at Heideggerian phenomenology through a more recent philosophical lens provided by McDowell (1994).[4]

McDowell draws a contrast between explanations at the personal level, and those that proceed subpersonally. So as not to be distracted by issues of personhood, I shall refer to these different levels as the *agential* and the *subagential*. (McDowell himself applies the distinction to frogs, so no real damage is done here.) Agential explanations are concerned with the identification and clarification of the constitutive character of agency – roughly, what it is to competently inhabit a world – whereas subagential explanations are concerned with mapping out the inner states and mechanisms (the parts of agents, as it were) that causally underpin agential-level phenomena. As the description of Heideggerian phenomenology given just above makes clear, that programme is a species of agential explanation. Cognitive science, by contrast, is in the business of supplying subagential explanations. So, in order to appreciate what phenomenology might tell cognitive science, we need to understand the relationship between agential and subagential explanations.

If our two explanatory levels were wholly independent of each other, the psychologist who is sceptical about what she can learn from phenomenology would have a point. Fortunately, however, McDowell shows that this cannot be the case. On the one hand, the ultimate goal of cognitive science is to map out the subagential elements whose organization, operation and interaction make it intelligible to us how it is that unmysterious causal processes (such as those realized in brains) can give rise to agential phenomena. Given that what those agential phenomena are (their constitutive character) is a matter that may be pursued by philosophical reflection of one kind or another, there is a sense in which agential explanation isolates the phenomena that subagential science tries to explain. On the other hand, what cognitive science discovers at the subagential level may sometimes lead us to revise our conception of what the agential level phenomena are. What we have, then, is a process of mutual constraining influence between the two styles of explanation, a process that McDowell (1994, p.197) describes as a 'perfectly intelligible interplay'.

Just how this interplay works – what the inter-level constraints are exactly – is a non-trivial matter that McDowell doesn't really explore, beyond the important point that such an interplay can surely be given shape without requiring a reduction of the agential to the subagential. A proper account is beyond the scope of the present discussion too, but a few words will provide some sort of mandate for what follows.

---

[4]As it happens there is good textual evidence that Heidegger already had in play something like the McDowellian view that I am about to sketch. For discussion see Wheeler (2005a, chapter 5).

The Problem of Representation 325

Consider: it would be mysterious just why circumspection (the experiential character of smooth coping) contains no subjects and no objects, if what is actually chugging away 'underneath' that experience were a heavy-duty system of internally located representation-based reasoning algorithms that generate planned sequences of movements by drawing inferences from a detailed perceptual representation of an external world of objects. The mechanistic story would be, to borrow a phrase from McDowell, phenomenologically off-key. More generally, then, if there is to be an intelligible connection between subagential and agential explanations of cognition, we might often expect there to be certain general structural isomorphisms between any agential account and its subagential counterpart. At the very least it seems that such structural isomorphisms, where they exist, are sufficient for inter-level intelligibility, even if they aren't strictly necessary. With all due caution, then, it seems that we might use the results of disciplined phenomenological analysis as a defeasible guide to the structure of the underlying causal processing, and vice versa, with the aim of bringing the two into an illuminating equilibrium. So is there any evidence that the tripartite Heideggerian analysis described above, and the position on representations and representational warrant that it embodies, is one half of just such an equilibrium?

Action-Oriented Representation

Let's begin with detached theoretical reflection – the domain of presence-at-hand. There is a dispute about the relationship between detached reflection and online problem solving within a broadly Heideggerian framework that, so far, I have glossed over. The unresolved issue is this: is it, as I have argued here, that detached reflection is a radically different kind of agential cognitive achievement to online problem solving, suggesting, via the intelligibility constraint, the presence of two different species of underlying mechanism (Wheeler 2005a)? Or is it that the agential and subagential psychological processes responsible for offline reflection are in some way continuous with those responsible for online intelligence (Gallagher 2007)? How this debate turns out will determine whether the domain of presence-at-hand involves representations that are fully action-neutral, specifiable without necessary reference to the representing agent, and essentially context-independent (Heidegger's view as I have interpreted him), or representations that in some way retain the marks of action-specificity, egocentricity, and context dependence. If the former option turns out to be right, and I think it will, then it seems that orthodox cognitive science already gives us good models for the underlying representational capacities in question. Consider, for example, Marr's (1982) influential theory of vision. Marr assumes that the main task confronting vision is to derive representations of the 3D shapes of objects from 2D arrays of light intensity values at the retina, via intermediate viewer-centred representations. The output representations of Marrian vision are thus well-poised to be the raw material for detached theoretical reflection, as the Heidegger of my interpretation conceives it: the 3D shape of

an object is *not a property defined relative to any particular action*; one of the sub-tasks of vision is precisely to *eliminate agent-relativity* (viewer-centredness) in favour of agent-neutral properties; and models of the 3D shapes of objects are *context-independent* representations.

Now, what about the agential-level representations that, on my Heideggerian view, support online practical problem solving? Do these find isomorphic subagential counterparts anywhere in cognitive science? The answer, I think, is yes, most notably where embodied-embedded theorizing meets recent work in robotics. In order to build complete robots that are capable of integrating perception and action in real time so as to generate fast and fluid embodied adaptive behaviour, various thinkers (e.g. Agre and Chapman 1990; Brooks 1991; Mataric 1991; Franceschini et al. 1992) have pursued an idea that I shall call *action-oriented representation*.[5] A seminal example, and one that illustrates the Heideggerian connection that I wish to forge, is provided by Mataric's sonar-driven mobile robot Toto (Mataric 1991). Toto wanders around its office environment following walls and avoiding obstacles. As it proceeds it constructs an internal map based on landmarks and that map then enables it to navigate between locations. So far this all sounds rather pedestrian. However, what is theoretically interesting about Toto's maps is that navigation-related information is encoded in them as patterns of sensorimotor activity. For example, if Toto keeps detecting proximally located objects on its right hand side, while its compass bearing remains unchanged, then a 'right-wall' is encoded in the subagential map, not as some objective entity, but in terms of the robot's sensorimotor 'experience' at the time. These structured sensorimotor 'experiences' (Toto's landmarks) are stored as connected nodes in a distributed graph, and this record of the robot's own sensorimotor history constitutes its subagential map of the spatial environment. Toto is then able to navigate its way around using paths encoded as sequences of past, current, and expected sensorimotor 'experiences'.[6]

It is now possible to see how the profile of Toto's subagential representations reflects that of the agential-level representations that populate the phenomenal experience of online practical problem solving. Toto's maps are: (i) *action-specific*, in that they are tailored to the job of producing the specific behaviour required; (ii) *egocentric*, in that they encode the environment in terms of the robot's own history of sensorimotor 'experiences'; and (iii) *intrinsically context-dependent*, in that context is woven into the representation-using mechanism's basic operating principles. It is worth saying a little more about (iii) which, in the case of Toto anyway,

---

[5] I have borrowed the term 'action-oriented representation' from Clark (1997) who introduced it in order to capture a notion of representation-in-action that is similar to the one that I am about to unpack, although minus the Heideggerian overtones.

[6] It is arguable that the manner in which Toto's action-oriented representations figure in the robot's route-planning processes qualifies those states as *decoupleable* (although of course not always *decoupled*) surrogates for environmental factors. However, not all states that manifestly count as action-oriented representations are decoupleable in this way, which is why decoupleability cannot be necessary for representation. For discussion, see Wheeler (2005a, chapter 8).

might be seen as a consequence of (i). Toto's maps encode spatial information as patterns of sensorimotor activity. This is a largely task-specific solution, appropriate for the navigational context. Such structures presumably wouldn't be much good for a vast range of other purposes, such as ordering correctly-sized carpets for the corridors or determining the precise distance to the snack bar. Because of their task-specific design, however, Toto's subagential mechanisms never confront the problem of selecting, out of a vast sea of potentially available spatial information that could have been retrieved and internally stored, the subset of such data relevant to the navigational context in which they operate. As one might put it, they don't confront the problem of finding their way into the right context. Why not? Because context is implicitly embedded in the mechanisms' basic operating principles. This is an issue to which we shall return below.

The claim on the table is that embodied-embedded cognitive science invites an approach to the question of representational warrant that in some interesting way recapitulates Heideggerian phenomenology. Thus we should expect there to be embodied-embedded models in which the orthodox commitment to the representational explanation of intelligent action is rejected in favour of an underlying causal story that arguably makes intelligible the nonrepresentational structure of smooth coping. And, given the context-embeddedness of smooth coping, we should expect the mechanisms postulated by such models to bear the mark of intrinsic context-dependence. Here is one model that fits the bill. At any one time animals (including human beings) do one thing rather than another, and what this is changes as circumstances change. This is known as the action selection problem. From a phenomenological perspective, however, it seems clear that although we may talk about action selection as a problem posed to the agent, the transition from one action to another is often a matter of smooth coping rather than practical problem solving strictly conceived. Unfortunately this clashes unhappily with the standard approach in robotics which assumes the subagential internal representation of appropriate behaviours and some sort of internal arbitration mechanism to decide between them. Adopting an embodied-embedded approach, Seth (1998) shows that, in a simple artificial world of power sources and traps, action selection desiderata such as prioritising with respect to currently relevant needs, sequencing behaviours appropriately, and opportunistic behaviour change, can be achieved by a minimal wheeled robot control architecture in which there are no internal representations of behaviour. Rather, a suite of independent artificially evolved activation functions directly link sensing and movement. The outputs (movement 'recommendations') from these sensorimotor connections are numbers that are simply combined (roughly, summed and scaled) at the wheels as part of an ongoing perception-action cycle. Plausibly this constitutes a non-representational mechanism for smooth coping, although of course much would need to be said about how such non-representational mechanisms might be integrated with mechanisms featuring action-oriented representations, within a single sub-agential architecture that undergoes transitions between smooth coping and online practical problem solving.

What exactly distinguishes action-generating mechanisms featuring action-oriented representations from those that don't? One answer here appeals to structural

isomorphisms between the agential and subagential levels, such that the representational interpretation of the subagential mechanism will be mandated by the representational phenomenology at the agential level. This might be one strategy, but is there anything one can say purely at the subagential level? The answer, I think, is positive: there are architectural factors that provide necessary and sufficient conditions for a target inner state to be accorded the status of subagential representation, by showing how the standing-in-for relation (see above) may be established at that level. These factors are: (a) being a genuine source of adaptive richness and flexibility (requiring that the inner state be causally correlated upstream with objects and states of affairs, and downstream with behavioural outcomes); (b) arbitrariness (interpreted as multiple realizability in which the class of equivalent mechanisms is fixed by informational rather than first-order-physical factors); and (c) being part of an homuncular system (a set of hierarchically organized, communicating modules, each of which performs a well-defined sub-task that contributes towards the collective achievement of an adaptive solution). There is much to be said about the arguments for and the details of this picture (see, e.g. Wheeler and Clark 1999; Wheeler 2005a), but it is surely plausible that the inner economy it describes warrants a representational gloss according to which the homuncular modules use the information-bearing elements to *stand in for* worldly factors (objects and states of affairs) in their communicative dealings.

From what we have seen so far, there seems to be no deep tension in the vicinity of cognitive science between representationalism and the (Heideggerian) phenomenological perspective. Of course, the scope of representational explanation turns out to be more limited than some orthodox cognitive scientists might have supposed, since phenomenological analysis suggests that the skilled adaptive intelligence characteristic of smooth coping does not reward a representational treatment; and the form that representations might take now has an additional dimension, given the action-oriented representations that figure in online problem solving. But so far what we have witnessed is a rethinking rather than a sweeping away. The tide, however, is about to turn, because there is a phenomenologically driven argument to suggest that it is precisely in virtue of adopting a representational account of intelligent action – any representational account that is, including an action-oriented one – that cognitive science opens the door to one of the most serious unsolved problems that it faces.

The Frame Problem

The frame problem[7] is located by the following question: how is it that an intelligent agent, conceived as a purely mechanistic system, is able to home in on just those aspects of all the things it senses, knows, or believes that are relevant to the present context of activity, while ignoring everything that is contextually irrelevant, and

---

[7]For introductions to the frame problem, see Dennett (1984), Pylyshyn (1987), Shanahan (2008).

how is it that that agent is then able to revise or act on that information in a contextually appropriate manner? In fact there are two different kinds of context-sensitive relevance here which may be identified with two different dimensions to the frame problem (Wheeler 2008). The first, call it the *intra-context frame problem*, challenges us to say how a purely mechanistic system is able to achieve appropriate flexible and fluid action *within* a context. The second, call it the *inter-context frame problem* challenges us to say how any purely mechanistic system is able to flexibly and fluidly switch *between* contexts in a relevance-sensitive manner.

To bring the frame problem into sharper focus, consider Hubert Dreyfus's description of the plight of an AI-programmed computer confronted by incoming environmental data:

> The significance to be given to each logical element [each internally represented piece of data] depends on other logical elements, so that in order to be recognized as forming patterns and ultimately forming objects and meaningful utterances each input must be related to other inputs by rules. But the elements are subject to several interpretations according to different rules and which rule to apply depends on the context. For a computer, however, the context itself can only be recognized according to a rule…

> [T]o pick out two dots in a picture as eyes one must have already recognized the context as a face. To recognize this context as a face one must have distinguished its relevant features such as shape and hair from the shadows and highlights, and these, in turn, can be picked out as relevant only in a broader context, for example, a domestic situation in which the program can expect to find faces. This context too will have to be recognized by its relevant features, as social rather than, say, meteorological, so that the program selects as significant the people rather than the clouds. But if each context can be recognized only in terms of features selected as relevant and interpreted in terms of a broader context, the AI worker is faced with a regress of contexts. (Dreyfus 1992, pp.288–289)

Dreyfus's parable brings into view the full horror of the frame problem. An obvious computational response to the challenge of selecting appropriate (sensory or stored) information is to equip the system with internally stored relevancy heuristics (represented rules of thumb) or representations of context. But all this does, unfortunately, is push the real problem one stage back. For how does the system decide which of its stored heuristics or potentially context-specifying representations are relevant? Another, higher-order set of heuristics or representations would seem to be required. But of course the same issue will re-emerge at that higher level. This regress is one signature of the frame problem.

Drawing on Heidegger and Merleau-Ponty, Dreyfus argues that one source of the frame problem is the orthodox cognitive-scientific assumption that intelligent action is fundamentally representation-driven. As he puts it, 'for Heidegger, all representational accounts are part of the problem' (Dreyfus 2008, p. 358). Dreyfus's account of the link between representation and the frame problem has three strands.[8] First he points out that phenomenological analysis reveals contextual significance to be a large-scale network of semantic connections whose massively holistic character

---

[8] Here I shall give just a brief summary of this analysis. For the details see Dreyfus (1990). For discussion see Wheeler (2005a, 2008).

renders it resistant to specification in terms of determinate representational content. Secondly he argues that because sensitivity to context-embedded relevance is a kind of knowing-how-to-navigate-one's-world, any attempt to representationally encode that knowledge is bound to generate difficulties: representations paradigmatically realize a form of knowledge-that-something-is-the-case, but knowledge-how (he suggests) cannot be reduced to knowledge-that. Finally he claims that representations are intrinsically context-independent structures to which context-dependent significance must somehow be added (see the earlier account of the representations that figure in Marrian vision); so any attempt to add significance by adding representations must result in a regress, since those additional representations will themselves need to have significance added by yet further representations (and so on).

So how do we avoid the frame problem? If, as Dreyfus claims, representations are part of the problem, then a giant step in the right direction will be taken if we can construct a nonrepresentational account of how intelligent action achieves its distinctive sensitivity to context-dependent relevance. And, it seems, phenomenology might be called upon to provide just such an account. Here Dreyfus (2008, p.340) draws on Merleau-Ponty's (1962) notion of the *intentional arc*.

> According to Merleau-Ponty, as an agent acquires skills, those skills are "stored," not as representations in the agent's mind, but as the solicitations of situations in the world. What the learner acquires through experience is not represented at all but is presented to the learner as more and more finely discriminated situations. If the situation does not clearly solicit a single response or if the response does not produce a satisfactory result, the learner is led to further refine his discriminations, which, in turn, solicit ever more refined responses. For example, what we have learned from our experience of finding our way around in a city is "sedimented" in how that city looks to us. Merleau-Ponty calls this feedback loop between the embodied coper and the perceptual world the intentional arc.

So, to adapt an example from Gallagher (2008), when poised to engage in the action of climbing a mountain, I do not build a representation of the mountain and infer from that plus additionally represented knowledge of my own abilities that it is climbable by me. Rather, from a certain distance, in particular visual conditions, the mountain 'simply' looks climbable to me. My embodied know-how in the domain of climbing is "sedimented" in how the mountain looks to me.[9] This idea dovetails with Heidegger's notion of *thrownness*, according to which the agent is in primary epistemic contact not with bare context-independent elements, but rather with equipment, the kind of entity that comes already laden with context-dependent

---

[9]Two points to note: First, as Gallagher points out, this model does not imply that perception in action is infallible. From a certain distance, the mountain may look climbable to me, but once I actually attempt to scale it I may discover that I was wrong. When I change my body-mountain relation, what the mountain affords may change too. Secondly, the model ultimately needs to support (or explain away) the observation that I may prepare myself for an ascent by reading mountaineering books and by looking at maps of the relevant mountain, such that presumably I form offline representations of various aspects of the problem domain that somehow feed into, or have a role to play in priming, the nonrepresentational perceptual experience of the mountain looking climbable to me. Thanks to Daniel Schmicking for bringing this second issue to my attention.

significance. Since human agents are characterized by thrownness, they are always already embedded in some meaningful context, so they are never in the position of having to add contextual significance to context-independent primitives (see e.g. Dreyfus, 1992, pp.262–263). For a representation-shunning thrown agent, then, the frame problem doesn't arise.

It won't have escaped notice that this nonrepresentational account sits unhappily alongside the earlier discussion of action-oriented representation. Since both pictures seem to be driven by disciplined phenomenology, what can be said to relieve the tension? We can begin by noting the way in which *Being and Time* progresses from (what we might call) a descriptive phenomenology of everyday encounters with entities to (what we might call) a deep-structure phenomenological analysis designed to reveal the hidden a priori transcendental conditions of experience. The categories of nonrepresentational smooth coping (readiness-to-hand) and agential level action-oriented representation (un-readiness-to-hand) are aspects of the former, essentially preliminary investigation. What Dreyfus brings into view is the landscape disclosed by the latter analysis, a landscape that is ontologically more basic than smooth coping or action-oriented representational problem solving. The ontologically pivotal phenomenon that this reveals is what Dreyfus (2008) calls *background coping* or *ground-level intelligence*, a nonrepresentational knowledge of how to get around one's world that underpins both smooth coping and action-oriented representational problem solving. It is at this level that the skilled climber's embodied know-how opens up the world as a familiar place of climbable mountains.

For Dreyfus, it is the character of background coping that dissolves the frame problem, in both its intra-context and inter-context forms. On the basis of our constantly honed background know-how, we respond directly to relevance, with context-bound entities soliciting or summoning us to act in ways shaped by our past experiences. And the capacity for flexible context-switching that lies at the heart of the inter-context frame problem is explained by the fact that I can be summoned not only by the present situation, but also by other situations that, because they have been relevant in the past, lie on the horizon of my experience (Dreyfus 2008, p. 359).

As the McDowellian framework introduced earlier indicates, our understanding of what background coping might be is incomplete until we have a cognitive-scientific model of the causal mechanisms that render that phenomenon an intelligible aspect of the material world. Earlier we encountered a suggestive model due to Seth (1998) in which aspects of this problem – such as prioritising with respect to currently relevant needs, sequencing behaviours appropriately, and opportunistic behaviour change – are addressed by a minimal robot control architecture involving a set of independent activation functions that directly link sensing to movement by way of a simple mathematical calculation that combines the outputs of these functions at the point of action. However, once we confront the full complexity of human background coping, such minimal solutions will surely take us only a small part of the way. So what might take us further? When Dreyfus himself accepts the challenge to supply a mechanistic account of background coping (Dreyfus 2008,

pp. 347–357), he discusses at length the work of neuroscientist Walter Freeman (e.g. Freeman, 2000). What emerges from this discussion is a vision of the brain as a nonrepresentational dynamical system primed by past experience to actively pick up and enrich significance, a system whose constantly shifting attractor landscape is identified as physically grounding Merleau-Ponty's intentional arc by causally explaining how newly encountered significances change the whole perceptual world of the agent.

Without dwelling on the fine-grained details of the mechanisms envisioned by Freeman, we can appreciate their nonrepresentational character by noting that they are just one realization of the phenomenon that Clark (1997) has dubbed *continuous reciprocal causation*. This is causation that involves multiple simultaneous interactions and complex dynamic feedback loops, such that (a) the causal contribution of each systemic component partially determines, and is partially determined by, the causal contributions of large numbers of other systemic components, and, moreover, (b) those contributions may change radically over time. One way in which continuous reciprocal causation undermines representational explanation is by undermining modular explanation (Wheeler 2005a, b). In systems that exhibit continuous reciprocal causation, the performance of any particular sub-task will be underpinned by a large and shifting number of interacting components whose contributions are changing in highly complex ways. Thus it becomes prohibitively difficult and explanatorily unhelpful to attempt to specify distinct and robust causal-functional roles played by reliably reidentifiable parts of the system. Because of this, there will be no useful modular decomposition of the system. The inference from a lack of modules to a lack of representations is justified by the fact that homuncularity, identified earlier as a necessary condition for representation, is a form of modularity, so if modularity fails, so does homuncularity, and so does representational explanation.

We can now offer the following gloss on Dreyfus's analysis of the frame problem. If (a) background coping dissolves the frame problem, (b) the causal machinery underlying background coping is characterized by continuous reciprocal causation, and (c) continuous reciprocal causation is a nonrepresentational mechanism, then (d) the cognitive-scientific answer to the frame problem tracks the phenomenological one by being nonrepresentational in form. In my view this argument is on the right track, but a nagging worry remains. Continuous reciprocal causation plausibly bestows a certain kind of large-scale flexibility on a system, a flexibility ripe to account, in part, for the fluid context-switching highlighted by the inter-context frame problem. However, the fact that a machine may flexibly reconfigure itself on the basis of continuous reciprocal causation among its elements does not guarantee that the behaviours generated by that machine will remain contextually relevant. All that is assured is that the machine supports the kind of plasticity that, when harnessed appropriately (i.e. in context-sensitive ways), may help to generate fluid context-switching.

If continuous reciprocal causation is not the whole answer here, what else is required? We learned earlier that the intra-context frame problem may be neutralized by the property of intrinsic context-dependence, as possessed by certain kinds of causal mechanism. It's the way in which context is woven into the basic operating

principles of such mechanisms that, in the intra-context case, dissolves the recalcitrant difficulty of assigning relevance.[10] But we are at present in the territory occupied by the inter-context frame problem, so what we need to understand is how it is possible for unmysterious causal machinery to realize a kind of *meta-level* intrinsic context-dependence. That is, we need to identify a property – if Dreyfus is right, a nonrepresentational one – that explains how systemic reconfigurations enable a machine to keep its responses and behaviour relevant across *transitions and change within an open-ended flow of shifting contexts*. One can glimpse a simple-to-state (although not so obviously simple-to-solve) version of this problem, if one imagines an agent entering a situation whose complexity places it beyond the reach of the sort of minimal action-selection solution deployed by Seth, and in which, due to the fact that what should be done is currently under-determined, more than one intrinsically context-dependent mechanism is poised to take charge of behaviour. Although heuristic tricks and ploys (e.g. always choose the more complex behaviour) might make some headway against this control problem in artificially constrained scenarios, such strategies are unlikely to generalize. What we are still missing, it seems, is the key to a general and more fundamental solution to the inter-context frame problem, akin to the one provided by the property of intrinsic context-dependence in the intra-context case.

Does Dreyfus's discussion give us any clue as to what might be going on? As noted earlier, at the phenomenological level Dreyfus takes the capacity for flexible context-switching to be explained by the fact that I can be summoned not only by the present situation, but also by other situations that have previously been relevant to me. So one might think that our missing ingredient is the causal-mechanistic explanation for this cross-contextual summoning. Dreyfus (2008, p. 360) writes:

> If Freeman is right, our sense of familiar-but-not-currently-fully-present aspects of what is currently ready-to-hand, as well as our sense of other potentially relevant familiar situations on the horizon of the current situation, might well be correlated with the fact that brain activity is not simply in one attractor basin at a time but is influenced by other attractor basins in the same landscape, as well as by other attractor landscapes which under what have previously been experienced as relevant conditions are ready to draw current brain activity into themselves. According to Freeman, what makes us open to the horizonal influence of other attractors is that the whole system of attractor landscapes collapses and is rebuilt with each new rabbit sniff [Freeman has worked extensively on rabbit olfaction], or in our case, presumably with each shift in our attention. And after each collapse, a new landscape may be formed on the basis of new significant stimuli – a landscape in which, thanks to past experiences, a different attractor is active. This presumably underlies our experience of being summoned.

What are we to make of this analysis? To me it is unclear whether the crucial reconfiguration of the neural attractor landscape is supposed to be (i) *caused by* the

---

[10]To be clear, what I have been calling the intrinsic context-dependence of certain behaviour-generating mechanisms is also recognized by Dreyfus. Thus, during his positive appraisal of Freeman's neuro-dynamic account, Dreyfus writes: "according to Freeman, in an active, hungry animal the output from the isolated detector cells triggers a cell assembly *already tuned to detect the relevant input* on the basis of past significant experience" (Dreyfus, 2008, pp. 350–351, my emphasis).

attentional shift (as might be suggested by the parallel with the rabbit sniff and the talk of a new landscape being formed "on the basis of new significant stimuli") or (ii) *the causal basis of* the attentional shift (as might be suggested by the thought that the attractors in the landscape determine what we attend to). Either way, however, there is a worry. If (i) is the correct interpretation, then the shift in attention itself remains unexplained. But at least sometimes that shift in attention is presumably governed by, and thus presupposes, our grip on the way in which our context of activity is changing. To that extent, then, the tabled suggestion begs the question. On the other hand, if (ii) is the correct interpretation, then it seems to me that we are still owed an explanation of how it is that, out of all the attractors in the pre-transition landscape that have been significant in the past, and that might have become active, it is the relevant one that is ultimately selected. Once again, it seems, the key question is being begged. If these worries are justified, then, so far at least, Dreyfus's attractive phenomenological account of how we escape the inter-context frame problem lacks an adequate mechanistic underpinning.

Concluding Remarks

In the wake of the frame problem, how do things stand with the concept of representation, in the dialogue between phenomenology and cognitive science over the explanation of intelligent action? As we have seen, the agential phenomenology of online practical problem solving has a representational character, and certain intrinsically context-dependent subagential mechanisms feature action-oriented representations. In both cases the representations are context-dependent in nature. Given that context-sensitivity is a critical property of intelligent action, this might be taken as evidence that representations are significant explanatory entities both for phenomenology and for cognitive science. But it is important to appreciate that the context-dependence of these representations is an inherited property. At the agential level it may be traced to the capacity for background coping. Of course background coping should not be thought of as some sort of distinct phenomenological 'module' to which smooth coping or practical problem solving may or may not be added. Rather, background coping is exhibited in, by being a structural precondition for, these context-embedded activities. At the subagential level the context-dependence of the representations in question is inherited, in a structurally similar manner, from the kind of mechanism in which the representations figure. Thus at neither explanatory level do the representations themselves account for the context-sensitivity of intelligent action.[11]

---

[11]Additional, although on its own inconclusive, evidence for this conclusion is supplied by the dual observations that nonrepresentational smooth coping displays sensitivity to context and that, as argued earlier, the nonrepresentational mechanisms that plausibly underlie smooth coping possess the property of intrinsic context-dependence.

For the fan of representations, the up-side of this situation is that action-oriented representations don't usher in the frame problem in the way that Dreyfus claims all representations do. Action-oriented representations inherit their context-embeddedness from the intrinsically context-dependent mechanisms in which they figure, which also means that there is no need for context to be accessed via the explicit internal representation of massively holistic networks of significance. Moreover, to the extent that what is being represented by action-oriented representations remains knowledge *that* the environment is thus and so, it is a thus and so that is encoded in action-specific and agent-relative terms. Such content makes sense only against the backdrop of intrinsic context-dependence provided by the rest of the mechanism. This deflects the thought that one way in which representational explanation runs into the frame problem is by illegitimately assimilating knowledge-how to knowledge-that. For the knowledge-that in question is not doing the frame-problem-related job of explaining how we are equipped with the capacity to be flexibly sensitive to what is relevant in a particular context. But this way of protecting representational explanation against Dreyfus's sceptical onslaught has a profound flip-side, since it means that *representations are neither aspects of the background coping that grounds our flexible sensitivity to context-dependent relevance (agential level), nor are they causally explanatory of it (subagential level)*. This is, I think, a striking result. Like all great dynasties, the age of representation may be rather less permanent, or at least less central to the way the world works, than its architects have imagined it to be.

References

Agre PE, Chapman D (1990) What are plans for? In: Maes P (ed) Designing autonomous agents: theory and practice from biology to engineering and back. MIT Press, Cambridge MA, pp 17–34

Brooks RA (1991) Intelligence without reason. In Proceedings of 12th international joint conference on artificial intelligence, Morgan Kauffman, San Mateo, CA, pp 569–595

Clark A (1997) Being there: putting brain, body, and world together again. MIT Press, Cambridge, MA

Cliff D, Noble J (1997) Knowledge-based vision and simple visual machines. Philos Trans R Soc: Biol Sci 352(1358):1165–1175

Cummins R (1996) Representations, targets and attitudes. MIT Press, Cambridge, MA

Dennett DC (1984) Cognitive wheels: the frame problem of AI. In: Hookway C (ed) Minds, machines and evolution: philosophical studies. Cambridge University Press, Cambridge, pp 129–151

Dreyfus HL (1990) Being-in-the-World: a commentary on Heidegger's being and time, division I. MIT Press, Cambridge, MA

Dreyfus HL (1992) What computers still can't do: a critique of artificial reason. MIT Press, Cambridge, MA

Dreyfus HL (2008) Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. In Husbands P, Holland O, Wheeler M (eds) The mechanical mind in history. MIT Press, Cambridge MA. 331–71 (A shortened version of this paper appears under the same title in Philosophical psychology 20/2 (2007), 247–268. I have worked with the full version of the text and page numbers cited refer to that version.)

Fodor JA, Pylyshyn Z (1988) Connectionism and cognitive architecture: a critical analysis. Cognition 28:3–71

Franceschini N, Pichon JM, Blanes C (1992) From insect vision to robot vision. Philos Trans R Soc, Ser B 337:283–294

Freeman W (2000) How brains make up their minds. Columbia University Press, New York

Gallagher S (2007) Review of Michael Wheeler's. 'Reconstructing the cognitive world: the next step'? Mind 116:792–796

Gallagher S (2008) Are minimal representations still representations? Int J Philos Stud 16(3):351–369, special issue on Situated cognition: perspectives from phenomenology and science. In: Ratcliffe M Gallagher S (eds)

Heidegger M (1927) Being and time (trans: Macquarrie J, Robinson E). Basil Blackwell, Oxford

Hinton GE, McClelland JL, Rumelhart DE (1986) Distributed representations. In Rumelhart DE, McClelland JL (eds) Parallel distributed processing: explorations in the microstructure of cognition, vol 1: Foundations. MIT Press, Cambridge MA, pp 77–109

Marr D (1982) Vision: a Computational Investigation into the human representation and processing of visual information. W.H. Freeman, New York

Mataric M (1991) Navigating with a rat brain: a neurobiologically inspired model for robot spatial representation. In: Meyer J-A, Wilson S (eds) From animals to animats: proceedings of the first international conference on simulation of adaptive behavior. MIT Press, Cambridge, MA, pp 169–175

McDowell J (1994) The content of perceptual experience. Philos Quarterly 44(175):190–205

Merleau-Ponty M (1962) Phenomenology of perception (trans: Smith C). Routledge and Kegan Paul, London

Newell A, Simon H (1976) Computer science as empirical enquiry: symbols and search. Communications of the Association for Computing Machinery, 19. Reprinted in: Boden MA (ed) (1990) The philosophy of artificial intelligence. Oxford University Press, Oxford, pp 105–132

Pylyshyn Z (ed) (1987) The robot's dilemma. Ablex, Norwood NJ

Rowlands M (2006) Body language: representation in action. MIT Press, Cambridge, MA

Seth A (1998) Evolving action selection and selective attention without actions, attention or selection. In: Pfeifer P, Blumberg B, Meyer J-A, Wilson S (eds) From animals to animats 5: proceedings of the fifth international conference on simulation of adaptive behavior. MIT Press, Cambridge, MA, pp 139–146

Shanahan, M. (2008) 'The Frame Problem' *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta(ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/frame-problem/>.

Smolensky, P. (1988) 'The Constituent Structure of Connectionist Mental States: A reply to Fodor and Pylyshyn', *Southern Journal of Philosophy*, 26, Supplement, 137-62.

Thelen E, Smith LB (1993) A dynamic systems approach to the development of cognition and action. MIT Press, Cambridge, MA

Vacariu G, Terhesiu D, Vacariu M (2001) Toward a very idea of representation. Synthese 129:275–295

Varela FJ, Thompson E, Rosch E (1991) The embodied mind: cognitive science and human experience. MIT Press, Cambridge, MA

Webb B (1994) Robotic experiments in cricket phonotaxis. In: Cliff D, Husbands P, Meyer J-A, Wilson S.W. (eds) From animals to animats 3: proceedings of the third international conference on simulation of adaptive behavior. MIT Press/Bradford Books, Cambridge MA, pp 45–54

Wheeler M (2005a) Reconstructing the cognitive world: the next step. MIT Press, Cambridge, MA

Wheeler M (2005b) Friends reunited? Evolutionary robotics and representational explanation. Artificial Life 11(1–3):215–231, special issue on Embodied and situated cognition, Almeida e Costa F, Mateus Rocha L (eds)

Wheeler M (2008) Cognition in context: phenomenology, situated robotics and the frame problem. Int J Philos Stud 16(3):323–349, special issue on Situated cognition: perspectives from phenomenology and science, Ratcliffe M, Gallagher S (eds)

Wheeler M, Clark A (1999) Genic representation: reconciling content and causal complexity. Brit J Philos Sci 50(1):103–135

Action and Agency

Thor Grünbaum

Action and Agency

In this chapter, we will be asking how the notions of intentional action and agency are related and discuss different ways of thinking about agents' experiences of agency. I will focus on agents' experience of agency as they engage in and execute their intentional actions.[1] A number of distinctions important to our theories about and experimenting with intentional agency will be presented, and arguments in favour of viewing the experience of agency as having a complex phenomenology will be given. Let us begin by asking how we should conceive of intentional action. It is no easy task to define what an intentional action is. A commonsensical conception would be to say that an intentional action is an action an agent is performing because she has some reason to do so. An intentional action would then, as Anscombe said, be the kind of behaviour to which the "Why-question" in a certain sense has application (Anscombe 2000: 11), namely, in the sense that requires a reason for acting as an answer.

If this is a good way to single out intentional actions, it means that an intentional action cannot be identified independently of the agent's perspective, her practical attitudes and practical reasoning – in short, her knowledge, beliefs, and practical outlook on the world. An agent's behaviour can be described in numerous ways: as a

T. Grünbaum
Department of Media, Cognition and Communication & Danish National Research Foundation:
Center for Subjectivity Research, University of Copenhagen

[1] I think it is fair to say that the classical phenomenologists to a certain degree have neglected to investigate the phenomenological structures of agents' experience of acting (their current acting). They seem to have been more interested in describing the phenomenology of deliberation and decision-making – that is, describing what leads up to the agent's acting. Thus, Pfänder (1963) investigated the motivational structures of decision-making; Schütz (1960, ch. 1, and 1951) investigated different types of motives for acting and explanatory practices; Sartre (1943) was interested in bodily experience as a source of agency and described our experience of freedom; and Ricoeur (1950) described the intentionality of practical thinking as a form of world-directedness and investigated ways in which practical thinking (deliberation, planning, willing) is constraint by abilities.

displacement of the body, as a specific type of reaction to a stimulus, as a piece of marvellous motor co-ordination, as determined by socio-economic status, upbringing, sex, or whatever. But if we are interested in the behaviour as something the agent is intentionally doing, then we have to understand it as something that makes sense from the agent's perspective. For the mouse in the kitchen, it makes sense to move in *that* direction since it can smell the cheese-odour as coming from *that* direction. Similarly, a person sits down on a bench intentionally because she feels tired and *this* bench *here* affords comfortable relaxation.

It is clear that not all aspects of an agent's behaviour are intentional – only those aspects the agent is aware of doing for a reason (a reason may be "Just for the fun of it" or "Because it feels right"). Notice that if it is the agent's practical perspective that rules whether or not some behaviour or some aspect of the behaviour is intentional, it will normally exclude the finer aspects of the agent's bodily movements from being intentional. As an agent I have a reason to pick up *this* pen, but I am not aware of pre-shaping my hand for any specific reason, or moving with a specific velocity, recruiting this specific amount of muscle force. These are all things I do, they are all aspects of my behaviour, but they are not aspects of my behaviour under which my behaviour appear as intentional; rather, we should say that they are involved in what I am intentionally doing in the sense that I engage in whatever bodily behaviour is needed to do what I have a reason to do – namely, to pick up the pen.[2]

With great ingenuity Anscombe investigated the conditions under which this specific sense of the why-question is refused application – that is, conditions under which we cannot say that the agent is doing something intentionally. She listed three situations in which the question is refused:

1. The situation in which the agent is not aware of executing the action that the why-question is ascribing to her ("Sorry, I didn't know I was blocking your view")
2. The situation in which the agent knows that the event the why-question is inquiring about has or is occurring; but she knows only by way of observation from the "outside" ("I can hear the bell is ringing; but I didn't know I was the one ringing it")
3. The situation in which she experiences her action as being involuntary, that is, outside her control in such a way that she could have no direct awareness of the causes of her behaviour ("I didn't mean to splash water on the table; someone gave me a push") (ibid., 25).

These are all situations in which we could not conceive of ourselves or others as behaving intentionally. If we turn Anscombe's inquiry around, we have three preliminary conditions that must be in place in order for us to grasp something as an intentional action:

---

[2] For a discussion and distinction of the intentional aspects of one's behaviour and the motor aspects, or between intentional action and movement, see (Anscombe 2000, §30; Taylor 1964, ch. 1–3; Hornsby 1997, ch. 7, 2005).

(a) The agent must be aware of some relevant aspects of her own action.
(b) The agent's awareness of her own action cannot be a form of self-observation from the outside.
(c) The agent must experience her own behaviour as something she controls; in other words, she must be acting with some "sense of agency".

This looks neat, but when you start to ask how these three conditions are related to each other, it soon becomes apparent that they cannot be kept apart. Crucial to the agent's awareness of her own action is her experience of agency, and her experience of agency is perhaps best spelled out in terms of awareness of certain features of her action – and this awareness is not observational.

Notions of Agency

There is no general consensus in recent literature about how to understand the notion of agency, neither in the philosophical,[3] nor in the psychological.[4] We can follow Davidson (1971) and conceive of agency as something that we ascribe to agents. To say of some individual that she is the agent that did or executed some action is to attribute agency to the individual for the behavioural event in question (including "worldly" consequences, such as "the breaking of the window"). This means that not only was she somehow causally responsible for the occurrence of the event, she also had certain conceptions of herself as the agent of the event, she had the right kind of practical attitudes (beliefs, desires, intentions), and she had the right kind of experiences of control, a "sense of agency". If we attribute agency to a person for the breaking of a window, we not only describe the person as causally responsible for the breaking of the window, we also describe her as having certain conceptions of herself as the agent of the breaking of the window, as having certain practical attitudes with regards to the breaking of the window, and as having certain practical experiences in the execution of the action of breaking the window (sense of agency, feeling of control).

One way to get a handle on the notion(s) of agency is by looking at cases where it breaks down or is missing. One type of breakdown of agency is found in cases where the agent cannot refrain from engaging herself in behaviour that she wholeheartedly thinks she should avoid. Drug addiction, compulsive disorders, and blind obedience to authorities might give us examples of this kind of breakdown. One influential way of describing such cases has been proposed by Frankfurt. According

[3]See, for example, (Davidson 1971; Frankfurt 1978; Searle 1983, ch. 3; Taylor 1985b; Smith 1988; Ginet 1990, ch. 1–2; Brewer 1993; Hornsby 1997, ch. 8; Gallagher 2000a; Mossel 2005; Hohwy 2007; Bayne 2007; Pacherie 2007).

[4]See, for example, (Libet 1985; Neisser 1993; Jeannerod 1999; Jeannerod and Pacherie 2004; Wegner and Wheatley 1999; Wegner 2003; Frith et al. 2000; Frith 2005; Marcel 2003; Tsakiris and Haggard 2005; Haggard 2005).

to Frankfurt, a drug addict who wholeheartedly identifies himself with the project of avoiding drugs can in certain situations feel compelled to take the drug. The addict ascribes high value to avoiding drugs, identifies himself with this choice and withdraws himself from the activity of taking the drug. As a result, when the addict takes his drug, he sees his action as being alien to him: he does not conceive of himself as the agent of it, he withdraws himself from it. As Frankfurt writes:

> It is in virtue of this identification and withdrawal, accomplished through the formation of second-order volition, that the unwilling addict may meaningfully make the analytically puzzling statements that the force moving him to take the drug is a force other than his own, and that it is not of his own free will but rather against his will that this force moves him to take it. (Frankfurt 1988: 18)

So, on one understanding of agency, the unwilling drug addict does not conceive of himself as the agent of his drug taking behaviour.[5] However, this seems to be an overly strong conception of agency. We would still describe the addict's taking the drug as intentional. He is painfully aware of reasons for acting as he does, he intends to do it (in one sense, and intends not to do it, in another), and actively attempts to carry out the action.[6] The addict does not see himself as being passively moved around by a hidden force or puppet master, rather he has the experience of not being able to control *himself*. In other words, Anscombe's why-question is still adequately applied in these cases: the unwilling addict is aware of acting for a reason. Thus, rather than understanding the situation of the unwilling addict as a complete breakdown of agency, we should perhaps understand it as a breakdown of self-control or autonomy.[7]

To bring out this distinction, compare these cases of breakdown in self-control with a different type of breakdown of agency. Take the fascinating case of "anarchic hand syndrome" which inflicts some patients with frontal lobe lesions. In this syndrome, a person's hand may perform apparently goal-directed and purposive movements, like grasping food from a plate on a neighbouring table in a restaurant. It may look like an intentional action, but in fact it is not. The person is not in control of the movements of her hand. She can only stop its movement by brute force, say, by grasping the anarchic hand with the other hand.[8] This kind of case seems to present us with a case of complete loss of agency for the behavioural event. The behaviour is not something the agent is doing for a reason and the agent does not

---

[5] For a similar conception of agency, see Velleman (1992) and Taylor (1985a).

[6] For an emphasis on the active aspect of addictive behaviour, see Ainslie (2001) and Elster (2000, ch. 5).

[7] For a similar conclusion, see Mele (2003, ch. 10).

[8] See Della Sala et al. (1994), Marchetti and Della Sala (1998). In the latter article they write: "The patients are aware of the bizarre and potentially hazardous behaviour of their hand but cannot inhibit it. They often refer to the feeling that one of their hands behaves as if it has a will of its own, but never deny that this capricious hand is part of their own body" (p. 196). And later: "(The patients) are always well aware of their odd behaviour and consciously try to overrule the unwanted action by appeasing the wayward hand" (p. 202). For a different and conflicting account, see Riddoch et al. (2000, esp. p. 607).

have any sense of bodily control. Grasping food from the neighbouring table is not intentional under any description.

Different notions of agency are thus in play; notions that pick out different levels of agency, so to speak. A higher-order level that concerns the way in which the agent conceives of herself as an agent and thinks about her own choices and actions; and a first-order level that concerns her actual deciding to do something and acting.[9] These are clearly different notions of agency in the sense that we can have one without the other as demonstrated by the case of the unwilling drug addict: He experiences a loss of higher-order agency – a loss of self-control – but nevertheless has a first-order experience of agency for the actions he executes when he seeks out the drug and takes it. The experience of the addict is not like the experience of an "anarchic hand". This latter kind of experience suffered by a person with an "anarchic hand syndrome" makes it clear that when persons lose their first-order experience of agency, they will no longer be able to conceive of themselves as agents in the strong (higher-order) sense. There will be no experiential-behavioural material on which to build the reflection of whether or not one is doing what one wants, or sees oneself as wanting. No aspect of their behaviour can be described as intentional and the Anscombean why-question is denied any application.

Experience of Agency

First-order experience of agency does not presuppose any ability of higher-order thinking – thoughts about oneself as an agent, thoughts about one's own thoughts, desires and actions. It probably does not presuppose any conceptual abilities. We could say that this minimal notion of agency and intentional action is applicable even to non-conceptual creatures: all it requires is that the agent – be it a mouse, an infant, or an adult – is aware of a reason for acting. The cat lying in front of one mouse hole, rather than another hole a few metres further away, does so for a reason: it saw the mouse run into this hole and not the other, and that is the reason why it is now waiting in front of this one. Seeing where the mouse ran is not a mere causal antecedent, but is crucially something that makes the cat's behaviour intelligible or meaningful from the cat's perspective.[10] Being aware of a *reason* for acting – which does not need to amount to more than awareness of a goal, or perhaps merely something the agent saw – thus seems to be a part of our experience of agency, if by this experience we mean the kind of agency we must minimally ascribe to an agent if

---

[9] For a similar distinction between higher-order agency and first-order agency, see Gallagher (2004) and Gallagher and Zahavi (2007), ch. 8, where the distinction is made primarily in a discussion of schizophrenic experiences of delusion of control.

[10] For a similar conception of reasons for action, see Dretske (2006a, sect. 4, and 2006b), and Anscombe (2000), sect. 47. For a recent discussion of different notions of reasons for action and their relation, see Hornsby (2008) and Steward (2009).

her behaviour is an intentional action. This is a very undemanding or "thin" notion of reason for acting. It only requires that the agent engages in a specific goal-directed behaviour because of what she is consciously experiencing or thinking: from the perspective of the agent, pursuing a specific goal makes sense or is intelligible because of what she experiences or thinks.[11]

Arguably, being aware of a reason for acting is not all there is to an experience of agency. Imagine a person suffering from the "anarchic hand syndrome", and imagine that just before her anarchic hand by itself embarks on picking something up from a stranger's plate, the person saw the plate and decided to pick up some food from it. When picking up the food, the person would be aware of a reason for acting (some delicious food right there), but she would nevertheless not be acting intentionally. The standard way to go from here is to understand such examples as showing that some causal conditions must be missing in our story so far. The person was aware of a reason for acting, but she was not acting *because* of this reason.[12] I do not think we should reject or disregard these causal conditions and constraints, but here I will focus on another, more phenomenological way we could proceed from this "anarchic hand"-version of a deviant causal scenario (where the agent is aware of a reason but her reason is not in the right way causing her to act).

An intuitive suggestion is that in the case of "anarchic hand"-behaviour, the agent is missing the sense of controlling or guiding her own movements. In the contemporary debates, the idea that something like a minimal phenomenal sense of guidance or bodily control is necessary was first introduced by Frankfurt (1978). In this seminal paper, Frankfurt argues that the fact that some behaviour is goal-directed or purposive is not enough for it to be intentional, since that would make many biological processes, where no agents are involved, into intentional actions. If an agent's behaviour is to be intentional, then, in addition to the behaviour being goal-directed, it is necessary that the agent has some experience or phenomenal sense of guiding her behaviour towards its goal. Frankfurt argues that this sense of guidance is exactly what is missing in the cases of causal deviancy (as in the imagined situation of the "anarchic hand").

The Primitivist Conception

This idea has been picked up by several authors. Ginet (1990) argues that on any occasion an agent does something intentionally, she has an "actish phenomenal quality". Any intentional action, be it a mental act (like mentally saying the French word *peu*)

---

[11]A minority of philosophers insist that *awareness of reasons* is not required for *intentional action*. Often they will argue that *acting for a reason* entails a high level of conceptual and reflective abilities. For one formulation of this position, see Wakefield and Dreyfus (1991) and Dreyfus (2006). I disagree. I operate with a less demanding notion of *reason for acting*.

[12]For the classic discussion, see Davidson (1963, 1973).

or a bodily intentional action, has, according to Ginet, a purely mental component, a mental act. In the case of mentally saying the French word *peu* the mental act is identical with the mental saying, whereas in the case of bodily intentional behaviour, the mental act is identical with the mental exertion of one's body, the *volition*. In any case, "The mental act has what we may call (for lack of better term) an *actish* phenomenal quality. This is an extremely familiar quality, recognizable in all mental action, whether it be mentally saying, forming an image, or willing to exert force with a part of one's body (p. 13)." According to Ginet's definition of intentional action, for something to be a bodily intentional action the agent must perform this inner act of willing with an *actish phenomenal quality*. This is a primitive quality without any inner articulation or structure. We can therefore not describe the quality in literal terms, but only pick it out demonstratively as "*that* quality there", and only creatures capable of having the experience will be able to recognize it (p. 22).

It is important to notice that, according to Ginet, this actish phenomenal quality, which is an aspect of our experience of agency that must be present to the agent if her behaviour is intentional, is a phenomenal quality that can exist even if no bodily action occurred – it is purely mental. "Our experience of our voluntary exertion is a *mental* process that is separate from – and could exist in the absence of – any bodily exertion (p. 28)." On this account, an agent's minimal sense of being active is a primitive (unstructured) *quale* that is completely independent of bodily movement or activity and the experience of movement and perception of its consequences.

One influential way to argue for this independence of the experience of agency makes use of an example introduced by James (1950, p. 490). James presents the case of Dr. Landry's patient. This patient was apparently completely deafferented in his arm and was thus unable to sense his own movement. Landry writes of the patient: "If, having the intention of executing a certain movement, *I prevent him*, he does not perceive it, and supposes the limb to have taken the position he intended to give it (quoted by James 1950, p. 490)." According to Searle (1983, p. 89), this proves that the sense of agency ("experience of acting") is independent of movement: the patient has no experience of movement (being deafferented and blindfolded) and he is in fact not moving, but he nevertheless believes that he has moved – so, he must have an experience as of acting, a sense of agency that is independent of the experience of moving.[13]

The conception of the sense of agency as a "motor and experience independent" primitive quale is consistent with a proposal made by Gallagher (2000a, 2000b). According to Gallagher's proposal, the sense of agency is the personal level result of certain sub-personal *pre-motor* computations. Gallagher uses a model of motor control developed by Wolpert and co-workers.[14] Central to this model is the idea of a comparator mechanism that is able to make use of a copy of the motor commands

---

[13] Notice that this argument could only hope to establish that sense of agency is independent of experience of movement (physiologically speaking, the patient's complete motor system is activated). For further discussions of this type of argument from total failure, see Grünbaum (2008).

[14] See Wolpert et al. (2001) and Wolpert et al. (2003).

to compute the most likely sensory outcome of the commanded movement and compare it with the intended movement. In Gallagher's words:

> a comparator mechanism operates as part of a non-conscious premotor or "forward model" that compares efference copy of motor commands with motor intentions and allows for rapid, automatic error corrections […]. This mechanism […] anticipates the sensory feedback from movement and underpins an online sense of self-agency that complements the ecological sense of self-ownership based on actual sensory feedback […]. If the forward model fails, or efference copy is not properly generated, sensory feedback may still produce a sense of ownership ('I am moving') but the sense of agency will be compromised ('I am not causing the movement'), even if the actual movement matches the intended movement (Gallagher 2000a).[15]

This framework opens up obvious possibilities for neuro-cognitive investigations into the neural basis of voluntary behaviour. It pin-points an experience that we are all assumed to have when we act voluntarily and that we can all report on and of which we can indicate the timing, and it suggests possible neural mechanisms that could be responsible for the experience in question.[16]

To sum up the picture thus far: In order for something I do to be an intentional action of mine, I must have an experience of agency for the behaviour and its relevant consequences. According to the present proposal my experience of agency can be spelled out as my awareness of some reason for acting – some goal or purpose – and a primitive phenomenal sense of agency, an "actish quale". We may wonder whether this primitivist understanding of sense of agency as an "actish quale" really gives us sufficient material to understand the experience of agency involved when we do something intentionally. According to Frankfurt (1978), what is needed for some behaviour to be intentional is that the agent has a conscious sense of guiding or controlling her behaviour towards her goal. If this is the kind of sense of agency we are looking for, it may seem as if the primitivist conception falls short of delivering sufficient material for an account.

The Complex Conception

We may contrast the primitivist conception of sense of agency with a conception that sees the sense of agency as a complex or structured experience.[17] Frankfurt's

---

[15] Although the quoted papers by Gallagher seem consistent with a primitivist conception of the sense of agency it is unclear to what degree they are actually committed to this conception. Gallagher more generally holds an embodied view of action, and in later work he endorses a more complex conception of agency (Gallagher 2005; Gallagher and Zahavi 2007, ch. 8).

[16] It's important that the experience is supposed to be reportable and timeable, because this makes possible the search for co-occurring brain processes accessible through imaging studies. This invitation to cognitive sciences has been accepted by a number of psychologist and neuroscientists; just to mention some, Chaminade and Decety (2002), Farrer and Frith (2002), Tsakiris and Haggard (2005).

[17] Recently, various authors have proposed different versions of a complex theory of the sense of agency, (Horgan et al. 2003; Proust 2000, 2003; Pacherie 2007; Bayne 2007; Grünbaum 2008).

idea is that in order for some behaviour to be an agent's intentional action it is not enough that the agent's behaviour is controlled by some goal or representation of some goal since that would include non-intentional biological processes (e.g. the immune defence system) and designed physical systems (e.g. thermostats). It must also be the case that the agent has an experience as of being the one guiding the behaviour towards her goal. If this is on the right track, the sense of agency is perhaps best rendered as the agent's sense of consciously controlling her own behaviour.

Two types of cases bring out the problems with the primitivist claim that the actish feel of initiating or causing one's movement is the same as having a sense of consciously controlling one's behaviour. The first type is provided by cases of fast object-oriented action that we find ourselves automatically completing even though we want to inhibit them. In order to test this situation experimentally, Pisella et al. (2000) designed an experiment where subjects were instructed to point to a green target and where, on some trials, subjects were instructed that if the target jumped they should stop their touching action. These trials were compared to trials where subjects were instructed to go ahead and touch the target. The result was that there was no significant difference in the corrective response between the two groups. Both groups corrected their movement in correspondence with the new target position almost equally fast. Thus, the subjects in the stop-trial could not help correcting their movement and touch the target. "After touching the displaced target, subjects of the location-stop group were aware of their mistakes and spontaneously expressed strong frustration. Irrepressible motor corrections were thus driven toward the new target location" (p. 730).

Pissella et al. present us with what looks like a situation where the agents feel a certain loss of control. The agent is actively initiating an action, but in the course of carrying it out, conditions for its execution are no longer fulfilled, and the agent wants to stop it. This makes it apparent that she cannot stop it, i.e., she has no control over her movement which is controlled by some non-conscious automatic mechanism. She feels she is forced to complete the action even though she actively tries to prevent it. But the action is out of her hands, she is no longer its full controlling agent. If this is a viable description of the scenario, then I think this example elicits the fact that part of what it is to feel in control is a sense of being able to guide one's action towards its successful completion. This involves the idea that if the agent becomes aware that her action is not on the right track, then she should be able to change or stop her action. In other words, I suggest that the notion of a minimal sense of control or agency involves some sense of one's conditional power to control: If the agent somehow registers that what she does is not what she intends to be doing, then she should be able to change or stop her action. This conditional structure of one's sense of control implies that the sense of control involves some form of continuous minimal sensory monitoring of one's behaviour.

The second type of case is that of blindsight. Perenin and Rossetti (1996) and Marcel (1998) investigated the ability of people with blindsight to process certain spatial stimuli presented to their blind field of vision. It was shown that the subjects processed information concerning the shape, orientation, and distance of the objects

presented in their blind field. The subjects were thus able to grasp objects that they claimed they were completely unable to see. According to Marcel (1998), subjects insisted that they were not aware of the objects, even though their success in grasping the objects was well above chance level. They also expressed surprise concerning their own abilities. In another task, Marcel presented a word in the subjects' blind field, showed them a list of words in their sighted field, and asked them to choose the word closest in meaning to the one presented in their blind field. Again the subjects claimed to be unaware of words presented in the blind field. Nevertheless, their success rate was above chance level. They were unable to give rational reasons for their choice of words, as one of the subjects said, she somehow "felt forced to choose one word" (p. 1574). In other words, it was not an experience of mere blind guesswork. It was more like the experience of being driven by a hidden force or unintelligible feeling.

What might this situation be like for the blindsighted subject? Imagine the following scenario. In complete darkness, you are told, "Catch this!", and you somehow skilfully catch the object. Each time somebody throws something at you from different directions you somehow manage to catch it. You have no idea of how you are doing it. You are aware of no sensory cues to act on, other than the verbal one. In a situation like this, you would intentionally initiate the action. Perhaps on hearing somebody telling you "Catch this!", you just start to move in some accidental direction. In that sense you would have a sense of actively initiating the movement. You would, I suggest, nevertheless experience a certain loss of control of behaviour. You would experience that your arm was forced to move in a specific direction and that some strange, hidden power forced your fingers to curve in a certain way, etc. By contrast, in a normal catching-situation, you have visual contact with the object and by attending to the object you consciously allow the object to control certain parameters of your act. Direction and force of the movement, the finger-shaping, the timing, etc, is perfectly intelligible to you, because you can see or attend to the object of your action. Perceiving the object of action is thus important to our normal sense of control in ordinary object-oriented intentional action.[18] I want to bring out two points from the discussion of these two examples.

The first example shows, I submit, that the experience of control is a temporally extended and teleological business. In a particular environment I have to control my behaviour in such a way that it reaches its goal. This involves a constant sensitivity to environmental and psychological changes that might influence how I want to proceed. If my behaviour and experience of controlling it is to be attuned to such an environment, I must have a continuous awareness of my acting such that it is open to my changing or stopping it. The sense of one's own agency can therefore not lie merely in the sense of active initiation of movement.

The second example shows us, I think, that the intelligibility to the agent of her motor action depends on perceptual access to one's environment. Even if we say that the agent does not explicitly control the finer adjustments of her motor behaviour

---

[18] For a related discussion of "blindsight-grasping" and agency, see Campbell (2003).

directed at a particular environment, she does understand her motor behaviour because the environment is present or at least available to her. If she has no perceptual contact with her environment, her motor behaviour, the ground level of her intentional engagement with the world, will appear to her as unintelligible, and its success as a matter of pure luck.

Taken together, such cases give us a good motive for arguing that the sense of agency understood as the sense of control, firstly, cannot be separated from the agent's bodily awareness and perception of objects of action and consequences of action, and, secondly, cannot be an unstructured, primitive quale. One way to conceive of the role of perception in the sense of agency or control would be this. Consider the following example: I am intentionally taking a drink from the glass in front of me. In grasping the glass in order to drink from it, I allow the glass to steer my action oriented at it. I do not control my action by attending to my movements and matching them with what I am reflectively aware of intending to do. Rather, I control my action on the object by attending to the object and allowing it to determine certain aspects of my behaviour.[19] By consciously selecting a certain spatial object as a target for my action, a certain schema for movement is also selected or determined (see Neisser 1993). That is to say that in selecting *that* thing in front of me as my target for action, a future schematic motor path is laid down as well as certain future consequences of the action. Certain experiential patterns are thus schematically specified. In choosing to act on *that* thing in this way, I expect my acting to involve certain proprioceptive and tactile sensations, and I expect certain changes in my visual field concerning both myself and the object I am acting on. That is, in accordance with the intentional goal of my object-oriented action, in acting I "passively" form a specific kind of experiential expectancy pattern involving different sense modalities. The sense of agency would then result from the formation and continuous fulfilment of these sensory expectancies.[20]

The claim is not that agents reflect on their intentions and derive sensory consequences which they go on to compare attentively to sensory feedback. Thus, to say that I experience the fulfilment of expected sensory consequences is not the same as saying that I am attentively monitoring or noticing that everything is happening as I expected. To say that I am aware of this fulfilment is rather to say that I am aware that nothing unexpected is happening. I am poised for the detection of any anomaly. This means that I would detect it immediately if something unexpected

---

[19] This conception of agency as object-involving is familiar from Merleau-Ponty. He sees object-oriented action as a kind of object-dependent reference (see especially 1945, p. 161) and describes how perceived objects of action control aspects of our intentional behaviour (see ibid, p. 154).

[20] In his lectures on ethics from 1925, Husserl presented a similar description of the experience of willing (as involving continuous sensory fulfilment), see Husserl (1988) and Melle (1992). Notice that this complex conception of agency remains consistent with Wolpert's model of motor control. See in particular Frith et al. (2000). In contrast to Gallagher's use of the model (in his 2000a), it could now be stressed that the sense of agency necessary for the applicability of Anscombe's why-question would involve not only a pre-motor comparator but also comparisons between expected sensory consequences and perceptual feedback. See also Hohwy and Frith (2004).

were to happen. If my movement in grasping the glass felt different from what I expected, if it felt strange in my hand, if it had an unexpected weight, or if the object behaved in a completely unexpected way as a consequence of my acting on it, then I would notice it. If the course of action did not evolve as expected, if it did not result in the expected sensory consequences and I were unable to adjust, I would experience a certain loss of control or agency. This continuous fulfilment experience grounds an experience of bodily control, i.e., it grounds the experience of having an implicit and conditional power not only to initiate, but also to alter and stop one's bodily activity.

Vision and Agency

On the primitivist conception, the sense of agency is a pure mental phenomenon that can exist independently of any actual movement, sensory feedback, and perception of one's surroundings. It is a qualia-theory of agency and the obvious place to look if you want to find the physical "supervenience base" of the sense of agency is in the brain. On the complex conception, by contrast, searching for a neural correlate of the sense of agency is no longer an obvious research strategy. The sense of agency is no longer conceived as the phenomenal aspect of the mental volition pre-dating the actual movement; rather, the sense of agency or control grows out of a continuous interplay between awareness of goal, awareness of corresponding sensory expectancy patterns, awareness of initiating one's bodily activity, perception of pertinent features or objects in one's surrounding selected by one's goal or intention, and the continuous sensory fulfilment of sensory expectations.

According to this complex conception, the sense of agency is thus a specific type of interplay between immediate aims, correlated sensory expectancies, and perception. This conception therefore breaks with a long and strong tradition in philosophy and psychology of conceiving of agency and volition as the exact opposite of perception. According to one version of this traditional view, perceptual experience delivers information that enables the subject to classify and identify the perceived objects, and to judge objects to be of this and that kind. Perceptual experience is thus thought to deliver the material for our belief formation and judgemental activity. As such, perception makes a contribution to our knowledge about the world and our long term memory. And importantly in this connection, in serving as input to this rational and context-independent, conceptual framework, perceptual experience is thought to make an important contribution to our practical reasoning and decision making – not directly, but in the form of beliefs and knowledge. According to such a picture, perceptual experience has no direct bearing on action, only indirectly by way of judgemental activity and practical reasoning.

In claiming that the sense of agency involves perception, we break with this kind of picture. The two cases of fast automatic error-correction and blindsight action not only suggest that continuous sensory fulfilment – and thereby monitoring – is important, but also that perceptual availability and selection of the object of action

is crucial to our ordinary sense of control. I shape, or rather I experience my bodily movements as shaped to fit the objects on which I am acting, and if the objects to which my actions are oriented were not perceptual available to me and I could not selectively attend to them, then the shape of my bodily experience would seem utterly unintelligible to me – as forced upon me and controlled by something of which I had no knowledge. In other words, I would experience my behaviour as being brought about involuntarily and the Anscombean why-question would no longer find any application.[21]

The perceptual availability of the surroundings and our ability to selectively attend to objects towards which we orient our behaviour is crucial to our experience of controlling what we are doing. Our conscious selective perceptual attention – the way we look at the world – is an integral part of our skilful behaviour. What one is looking at depends on what one intends to do and how skilled one is at doing it. To be skilled at doing something is in part knowing where one should be looking. This is why it is an important part of learning a new motor skill to learn which region of the world one ought to attend to: when you teach somebody to drive a car or a bicycle, you have to teach them not to attend to road just in front of them, but instead to look up and ahead, etc. The fact that conscious perceptual selection is part of ordinary skilful behaviour can also be seen in the way selection of objects and type (or degree) of skill depend on each other: an expert tennis player making a serve will most likely attend to the corner she wants the ball to hit or perhaps she is so skilled she can attend to the movements of her opponent. A novice player without any serving-skills will, by contrast, have to break the serve into a number of successive steps relying on other skills. The novice will thus break the serve into, for example, the actions of throwing the ball up in the air and trying to hit the ball (and hoping for the best). She will rely on her skill to throw something up in the air and she will first visually select the location in mid-air she wants the ball to reach and secondly, the ball she wants to hit. Selective attention is part of an agent's exercise of her behavioural skills and often we can determine what she is doing by determining what she is attending to.[22]

The structure of skilful behaviour and of the agent's experience of agency does consequently not support a strict separation of perceptual experience and action. Typically we visually control our intentional skilful behaviour – or at least visually control certain aspects of our behaviour – by our conscious selective attention. If our behaviour is to a large extent visually controlled, then it should come as no surprise that our sense of agency or control involves perception.[23]

---

[21] For a different but related critique of the classical "sandwich" model of perception-practical reasoning-action, see Hurley (1998).

[22] For a similar conception, see Merleau-Ponty's idea of an *intentional arch* (Merleau-Ponty 1945, part 1, 3, esp. pp. 155ff.).

[23] This has often enough been emphasized by psychologists working in the ecological and ideo-motor traditions, for example, Gibson (1977, 1979), Turvey (1977, 1992), Neisser (1985, 1993), Prinz (1992, 1997), Hommel et al. (2001).

Concluding Remarks: The Illusion of Agency

We started out by giving Anscombe's definition of intentional action as the kind of behaviour to which the why-question in a specific sense has application. The relevant sense is the one that asks for the agent's reason for engaging in the behaviour in question. We proceeded by arguing that in order for this question to find application, the agent must have an experience of agency. A distinction was then drawn between a higher-order conception of one's own desires, choices and intentional agency – a conception of agency as self-control – and a first-order experience of agency in the execution of intentional action. The application of the why-question – the question that asks for the agent's reason for her specific course of action – does not require a higher-order conception of agency, but only the first-order experience. One way to go from here is to say that this first-order experience of agency only needs to involve some awareness of one's aim in acting (awareness of reason) plus a primitive sense of agency, an *actish quale*. I labelled this the primitivist conception of the sense of agency.

The final question I want to raise concerns what it takes for this primitivist sense of agency to be satisfied or whether it can at all be satisfied. One natural way to conceive of satisfaction of an experience of agency is to say that one's experience is satisfied in the case where one is in reality active, is in reality performing a voluntary action. That is, if a person has an experience as of himself as active – even if she merely has a primitive actish feel – it is natural to ask if this experience or primitive feel is veridical or not: an actish feel is veridical in the case the agent is in fact acting intentionally.[24]

The fact that even the notion of a primitive actish phenomenal quality elicits this concern demonstrates that the primitive sense of agency is not supposed to be primitive in the sense of *non-representational* qualia.[25] The actish feel is supposed to signal activity; it is a mark of agency stamped onto some of my movements, and carries the information that these movements are actively initiated. If this phenomenal feel carries information, then it can misinform – it can be true or false.

Doubts can be raised, however, about whether it is at all possible for this kind of primitive phenomenal quality to be veridical. Can it at all carry information about true bodily activity? If we as many cognitive scientists conceive of the sense of agency as directly brought about by a neural mechanism, it seems that the answer must be no. Activity here means intentional activity, it is something agents do. But what room is left for the idea that agents do things intentionally when their experience of doing is brought about by some hidden neural mechanism? If the answer to this question is "none", then it seems that the sense of agency is not information about the fact that *I* am doing something actively – a sense that *I* am

---

[24] For a discussion of the issues of veridicality of the experience of agency, see Bayne and Levy (2006).

[25] For the conception of qualia as non-representational phenomenal qualities, see Shoemaker (1981), Block (1990), Chalmers (1996).

controlling my own behaviour – but rather is a phenomenal label glued onto some of my movements by my brain for some reason or other (that we can perhaps understand in terms of survival value or evolution). Our primitive sense of agency would no longer be an expression of our intentional activity but rather something that gives us an illusion of agency. Consequently, what we would need in terms of a psychological and a philosophical theory would be a way of explaining what brings about this illusory sense of agency – we would need an error-theory of agency.[26]

Some philosophers argue that this kind of error-theory of agency is an inconsistent position. We need to acknowledge that agents and their intentional agency are a metaphysical reality.[27] If this is true, we have another reason to be critical of the primitivist notion of agency. What we want is a notion of our experience of agency that can be veridical – an experience that can be true because we are truly active or in control. One reason for endorsing the version of the complex theory of the sense of agency I have proposed is that it can (presumably) accommodate this desideratum: on this account, the experience of agency grows out of the interplay between the agent's aim, her sensory expectancies, and her perception. It is the result of a specific kind of interplay between conscious features of the agent: the way she selectively attends to objects, how she is marginally aware of certain schematic sensory expectancies, and how she is aware, again marginally, of their continuous sensory fulfilment. It is an experience she has when she consciously directs her behaviour towards a goal, when she consciously engages in whatever behaviour is needed in order to reach her aim.

In this chapter I have drawn a number of useful distinctions but will hardly have succeeded in settling any important controversies. One thing is certain, though: if we want to understand the nature of intentional action, we need an account of the agent's experience of agency.[28]

References

Ainslie G (2001) Breakdown of the will. Cambridge University Press, Cambridge
Anscombe GEM (2000) Intention. Harvard University Press, Cambridge, MA
Bayne T (2006) Phenomenology and the feeling of doing: Wegner on the conscious will. In: Pockett S, Banks WP, Gallagher S (eds) Does consciousness cause behavior?. MIT Press, Cambridge, MA, pp 169–186
Bayne T (2008) The phenomenology of agency. Philos Compass 3: 1–21
Bayne T, Levy N (2006) The feeling of doing: deconstructing the phenomenology of agency. In: Sebanz N, Prinz W (eds) Disorders of Volition. MIT Press, Cambridge, MA, pp 49–68

---

[26] One prominent example of such an error-theory is Wegner (2002, 2005). For a critical philosophical discussion, see Bayne (2006).

[27] This is argued by Hornsby (2004).

Block N (1990. Inverted earth. In: Tomberlin J (ed) Philosophical perspectives, vol. 4. Ridgeview, Atascadero, CA, pp 52–79

Brewer B (1993) The integration of spatial vision and action. In: Eilan N, McCarthy R, Brewer B (eds) Spatial representation. Oxford University Press, Oxford, pp 294–316

Campbell J (2003) The role of demonstratives in action-explanation. In: Roessler J, Eilan N (eds) Agency and self-awareness. Oxford University Press, Oxford, pp 150–164

Chalmers D (1996) The conscious mind. Oxford University Press, Oxford

Chaminade T, Decety J (2002) Leader or follower? Involvement of the inferior parietal lobule in agency. NeuroReport 13(1528):1975–1978

Davidson D (1963) Actions, reasons, and causes. J Philos 60. Reprinted in his Essays on actions and events. Oxford University Press, Oxford 1980/2000, pp 3–20

Davidson D (1971) Agency. In: Binkley R, Bronaugh R, Marras A (eds) Agent, action, and reason. University of Toronto Press, Toronto. Reprinted in his Essays on actions and events, Oxford University Press, Oxford 1980/2000, pp 43–62

Davidson D (1973) Freedom to act. In: Honderich T (ed) Essays on freedom of action. Routledge and Kegan Paul, London. Reprinted in his Essays on actions and events, Oxford University Press, Oxford, 1980/2000, pp 63–82

Della Sala S, Marchetti C, Spinnler H (1994) The anarchic hand: a fronto-mesial sign. In: Boller F, Grafman J (eds) Handbook of neuropsychology. Elsevier, Amsterdam

Dretske F (2006) Perception without awareness. In: Gendler TS, Hawthorne J (eds) Perceptual experience. Oxford University Press, Oxford, pp 147–180

Dreyfus H (2006) Overcoming the myth of the mental. Topoi 25(1–2):43–49

Elster J (2000) Strong feelings. Emotion, addiction, and human behaviour. MIT Press, Cambridge, MA

Farrer C, Frith C (2002) Experiencing oneself vs. another person as being the cause of an action: the neural correlates of the experience of agency. NeuroImage 18:324–333

Frankfurt H (1978) The problem of action. Am Philos Q 15(2):157–162

Frankfurt H (1988) Freedom of the will and the concept of a person. In his The importance of what we care about. Cambridge, Cambridge University Press, pp 11–25

Frith C (2005) The self in action: lessons from delusions of control. Conscious Cogn 14:752–770

Frith C, Blakemore S-J, Wolpert DM (2000) Abnormalities in the awareness and control of action. Phil Trans R Soc Lond 355:1771–1788

Gallagher S (2000a) Philosophical conceptions of the self: implications for cognitive science. Trends Cogn Sci 4(1):14–21

Gallagher S (2000b) Self-reference and schizophrenia: a cognitive model of immunity to error through misidentification. In: Zahavi D (ed) Exploring the self: philosophical and psycho-pathological perspectives on self-experience. John Benjamins, Amsterdam/Philadelphia

Gallagher S (2004) Agency, ownership, and alien control in schizophrenia. In: Zahavi D, Grünbaum T, Parnas J (eds) The structure and development of self-consciousness. John Benjamins, Amsterdam, pp 89–104

Gallagher S (2005) How the body shapes the mind. Oxford University Press, Oxford

Gallagher S, Zahavi D (2007) The phenomenological mind. An introduction to philosophy of mind and cognitive science. Routledge, London

Gibson JJ (1977) The theory of affordances. In: Shaw R, Bransford J (eds) Perceiving acting, and knowing. Lawrence Erlbaum, Hillsdale, NJ, pp 67–82

Gibson JJ (1979) The ecological approach to visual perception. Lawrence Erlbaum, Hillsdale, NJ

Ginet C (1990) On action. Cambridge University Press, Cambridge

Grünbaum T (2008) Trying and the arguments from total failure. Philosophia Philosophic Qu Israel 36(1):67–86

Grünbaum T (2008) The body in action. Phenomenol Cogn Sci 7: 243–261

Haggard P (2005) Conscious intention and motor cognition. Trends Cogn Sci 9(6):290–295

Hohwy J (2007) The sense of self in the phenomenology of agency and perception. Psyche 13:1

Hohwy J, Frith C (2004) Can neuroscience explain consciousness? J Conscious Stud 11(7–8):180–198

Hommel B, Müsseler J, Aschersleben G, Prinz W (2001) The Theory of Event Coding (TEC): a framework for perception and action planning. Behav Brain Sci 24:849–937

Horgan T, Tienson J, Graham G (2003) The phenomenology of first-person agency. In: Walter S, Heckmann H-D (eds) Physicalism and mental causation: the metaphysics of mind and action. Imprint Academic, Exeter, pp 323–340

Hornsby J (1997) Simple mindedness. Harvard University Press, Cambridge, MA

Hornsby J (2004) Alienated agents. In: De Caro M, Macarthur D (eds) Naturalism in question. Harvard University Press, Cambridge, MA, pp 173–187

Hornsby J (2005) Semantic knowledge and practical knowledge. Proc Aristotelien Soc, Supplementary volume, 79

Hornsby J (2008) A disjunctive conception of acting for reasons. In: Haddock A, Macpherson F (eds) Disjunctivism: perception, action, knowledge. Oxford University Press, Oxford

Hurley S (1998) Consciousness in action. Harvard University Press, Cambridge, MA

Husserl E (1988) Vorlesungen über Ethik und Wertlehre 1908–1914, Husserliana XXVIII. Kluwer, Dordrecht

James W (1950) The principles of psychology, vol II. Dover, New York

Jeannerod M (1999) To act or not to act: perspectives on the representation of action. Q J Exp Psychol 52A:1–29

Jeannerod M, Pacherie E (2004) Agency, simulation and self-identification. Mind & Lang 19(2):113–146

Libet B (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action. Behav Brain Sci 8:529–566

Marcel A (1998) Blindsight and shape perception: deficit of visual consciousness or of visual function? Brain 121:1565–1588

Marcel A (2003) The sense of agency: awareness and ownership of action. In: Roessler J, Eilan N (eds) Agency and self-awareness. Oxford University Press, Oxford, pp 48–93

Marchetti C, Della Sala S (1998) Disentangling the alien and anarchic hand. Cogn Neuropsychiatry 3(3):191–207

Mele A (2003) Motivation and agency. Oxford University Press, Oxford

Melle U (1992) Husserls Phänomenologie des Willens. Tijdschrift voor filosofie 54:280–304

Merleau-Ponty M (1945) Phénoménologie de la perception. Gallimard, Paris

Mossel B (2005) Action, control and sensations of acting. Philos Stud 124:129–180

Neisser U (1985) The role of invariant structures in the control of movement. In: Frese M, Sabini J (eds) Goal directed behavior: the concept of action in psychology. Lawrence Erlbaum, Hillsdale, NJ, pp 97–109

Neisser U (1993) The self perceived. In: Neisser U (ed) The perceived self. University of Cambridge Press, Cambridge, pp 3–21

Pacherie E (2007) The sense of control and the sense of agency. Psyche 13:1

Perenin MT, Rossetti Y (1996) Residual grasping in a henianopic field: a further dissociation between perception and action. NeuroReport 7:793–799

Pfänder A (1900 and 1911/1963) Phänomenologie des Wollens & Motive und Motivation. Verlag Johann Ambrosius Barth, München

Pisella et al. (2000) An 'automatic pilot' for the hand in human posterior parietal cortex: toward reinterpreting optic ataxia. Nat Neurosci 3(7):729–736

Prinz W (1992) Why don't we perceive our brain states? Eur J Cogn Psychol 4:1–20

Prinz W (1997) Perception and action planning. Eur J Cogn Psychol 9(2):129–154

Proust J (2000) Awareness of agency: three levels of analysis. In: Metzinger T (ed) The neural correlates of consciousness. MIT Press, Cambridge, MA, pp 307–324

Proust J (2003) Perceiving intentions. In: Roessler J, Eilan N (eds) Agency and self-awareness. Oxford University Press, Oxford, pp 296–320

Ricoeur P (1950) Philosophie de la volonté I. Le volontaire et l'involontaire. Aubier, Paris

Riddoch MJ, Humphreys G, Edward MG (2000) Visual affordances and object selection. In: Monsell S, Driver J (eds) Control of cognitive processes (attention and performce XVIII). MIT Press, Cambridge, MA, pp 603–626

Sartre J-P (1943) L'être et le néant. Gallimard, Paris

Schutz A (1951) Choosing among projects of action. Philos Phenomenol Res 12:161–184

Schütz A (1932/1960) Der sinnhafte Aufbau der sozialen Welt. Springer Verlag, Wien

Searle J (1983) Intentionality. Cambridge University Press, Cambridge

Shoemaker S (1981) The inverted spectrum. J Phylos 74(7):357–381

Smith AD (1988) Agency and the essence of actions. Philos Q 38(153):401–421

Steward H (2009) Animal Agency. Inquiry 52(3): 217–231

Taylor C (1964) The explanation of behaviour. Routledge and Kegan Paul, London

Taylor C (1985a) "What is human agency?" in his *human agency and language: philosophical papers I*. Cambridge, Cambridge University Press, pp 15–44

Taylor C (1985b) Hegel's philosophy of mind. In his *human agency and language: philosophical papers I*. Cambridge University Press, Cambridge, pp 77–96

Tsakiris M, Haggard P (2005) Experimenting with the acting self. Cogn Neuropsychol 22(¾):387–407

Turvey MT (1977) Preliminaries to a theory of action with reference to vision. In: Shaw R, Bransford J (eds) Perceiving acting, and knowing. Lawrence Erlbaum Associates, Hillsdale, NJ, pp 211–265

Turvey MT (1992) Affordances and prospective control: an outline of the ontology. Ecol Psychol 4(3):173–187

Velleman JD (1992) What happens when someone acts. Mind 101(403):461–481

Wakefield J, Dreyfus H (1991) Intentionality and the phenomenology of action. In: Lepore E, Van Gulick R (eds) John Searle and his critics. Blackwell, Oxford, pp 259–270

Wegner D (2002) The illusion of conscious will. MIT Press, Cambridge, MA

Wegner D (2003) The mind's best trick: how we experience conscious will. Trends Cogn Sci 7(2):65–69

Wegner D (2005) Who is the controller of controlled processes? In Hassin R, Uleman JS, Bargh JA (eds) The new unconscious. Oxford University Press, New York, pp 19–36

Wegner D, Wheatley T (1999) Apparent mental causation: sources of the experience of will. Am Psychol 54(7):480–492

Wolpert DM, Grahramani Z, Flanagan JR (2001) Perspectives and problems in motor learning. Trends Cogn Sci 5(11):487–494

Wolpert DM, Doya K, Kawato M (2003) A unifying computational framework for motor control and social interaction. Phil Trans R Soc Lond B 358:593–602

Meaning, World and the Second Person[1]

Juan J. Botero

Theories of consciousness seem to raise some philosophical questions related to the good old issues concerning realism and anti-realism. Let's take the term "consciousness" in a very broad sense to signify the entire contentful "psychic" (mental) life. The problem I want to address is this: there seems to exist both serious empirical and sound theoretical arguments in support of the view that human reality, i.e., subjective conscious human experiences as well as the correlative "objective", "real", experienced human world, is subjectively constituted in conscious or brain operations; but how then is it that the *sense* of living in a world that exists independently of us, that is, a real, external world, is *constitutive* of our everyday conscious experience, and thus dictates a very strict objectivist stance from which to gain understanding about it? The empirical arguments are provided mainly by the neuroscientific study of consciousness. The more theoretical arguments can be found in classic phenomenology, more specifically Husserlian phenomenology, which is credited with uncovering, through a very strict method of philosophical inquiry, the subjective constitution of objective, transcendent reality. So then, these two independent approaches to consciousness seem to support both the view that the sense of the world is subjectively constituted, and the apparently contradictory view that this sense is the sense of an external, independent world. I will try to handle now this typical philosophical puzzle.

J.J. Botero
Departamento de Filosofía, Universidad Nacional de Colombia, Bogotá

So, this is what I am going to do:

First, I will outline a view on consciousness that, as far as I can tell, is supposed to be well-grounded and solidly supported by evidence and empirical work. This view appears to furnish evidence leading to the consequence that the world for us, the world we are conscious of in our everyday conscious life, is an upshot of the operations of the brain. For that reason I will call it an "internalist view on consciousness and cognition", which I deem is quite philosophically naïve. For this reason I will try to extract from it some philosophical consequences, mainly inspired by phenomenological considerations. After that, I will explore the outlines of an embodied, situational approach to conscious experience that seems well suited to overcome the criticisms that could be raised by this extension of the "internalist" view. Finally I will propose a not very substantial amendment of this latter approach in order to cope with the requirements of some phenomenological evidence. This part will be highly conjectural and speculative and is meant to suggest a possible path for further investigations.

The "Conscious Brain"

A particularly lucid formulation of the kind of approach I am interested in has been published by Rodolfo Llinás, a leading physiologist and neuroscientist, in his book *I of the Vortex* (Llinás 2001). But Llinás's explicitly reductionist account must be seen as just an example of what seems to be a more or less generalized standard approach to the phenomenon of consciousness in the field of the neurosciences. For some closely related views, see, for instance, Francis Crick (Crick 1994), and, more recently, Gerald Edelman (Edelman 2006).

One problem faced by much of current neuroscience and philosophy of mind is to explain how phenomena like brain processes (biological, chemical and electrical) produce phenomena like conscious experiences (feelings, sensations, but also meaning and intentionality). One way to address this problem is to ask what are brains and nervous systems and how do they actually operate to carry out their biological functions. According to Llinás, only multicellular creatures that have the biological property of motricity have nervous systems: "the evolutionary development of nervous systems is an exclusive property of actively moving creatures" (p. 17). Now, active movement in an external environment requires some kind of planning and guidance. This is what brains are for (p. 18). From here he comes to the hypothesis that

> The nervous system has evolved to provide a plan, one composed of goal-oriented, mostly short-lived predictions verified by moment-to moment sensory input (ibid.).

How does it work? Llinás expands his hypothesis by claiming that *prediction* is what provides such a plan. In his own words:

> The creature must anticipate the outcome of a given movement on the basis of incoming sensory stimuli. A change in his immediate environment must evoke a movement (or lack of it) in response to ensure survival. The capacity to predict the outcome of future events

–critical to successful movement– is, most likely, the ultimate and most common of all global brain functions. (p. 21)

To successfully perform its function of planning and guidance of active movement, there is a class of functional brain states that generate, by means of mechanisms we will consider later, what the author calls "sensorimotor images" of the external world, that is, "the conjunction or binding of all relevant sensory input to produce a discreet functional state that ultimately may result in action" (p. 1).

Before going further, it is worth noticing the apparently careless use of concepts such as "plan", "prediction", "verification", and "images" in a context where one would have expected to find descriptions delivered in a pure extensional, non-intentionalistic idiom. This is a quite pervasive characteristic one can find in the scientific accounts of mind and consciousness: many neuroscientists in their practice switch between these two domains of explanation all the time. They, as well as many philosophers working in a neuroscientific vein, use and manipulate neuro-physiological, chemical and physical principles and properties, while swiftly shifting to the use of concepts pertaining to a different domain of explanations. This category problem pervades the entire account I am trying to examine, and so it reflects the philosophical naivety I mentioned some lines earlier.

With this caveat in mind, let us continue with the theory. So far we have a sketch of an evolutionary hypothesis concerning the emergence of the brain, articulated in a good functional framework. Now, if that is what brains are for, what are they and how do they manage to perform their biological function? A basic, fundamental idea in this view is that a brain is "a living entity that generates well-defined electrical activity" (p. 2) and "operates as a closed system" (p. 8). These two features, intrinsic electrical activity and operating as a closed system, together with the role played by movement, constitute the backbone of the functional-physiological side of the theory. As far as I can tell, it is solidly supported by evidence and lab work.

The fact that the nervous system operates on its own, as a closed system, has the consequence that sensory inputs do not generate cognitive sates, for example perceptions. These are entirely generated by the workings of the nervous system, just "modulated" by sensory inputs:

> The *significance* of sensory cues is expressed mainly by their incorporation into larger, cognitive states or entities. In other words, sensory cues earn *representation* via their impact upon the pre-existing functional disposition of the brain (p. 8, emphasis added).

How does the brain control the body? Llinás traces the doings of the nervous system back to the cellular level, showing how the operation of the whole-brain derives its properties from the special properties of the cellular, neuronal level. The most relevant property of neurons involved in cognition and consciousness is intrinsic oscillation. This intrinsic activity at the level of neurons, amplified and phased out at a level of neuronal groupings, is, according to Llinás, "the foundation of the self" (see p. 9). For example, prediction, which is pervasive throughout practically all levels of brain functions, "begins at the single neuronal level" (p. 25). As for movement, the oscillatory neuronal behaviour is believed "to synchronize the elements of the motor apparatus as a whole to facilitate the combination of the different premotor

signals required for the generation of a meaningful movement" (p. 44). This oscillatory, periodic behaviour is the core of the timing device required to phase together the varying conduction velocities of the different neuronal pathways that must exercise control over movement. Studies strongly suggest that this synchronizing control system is located, or at least closely linked to, the operation of the inferior olive (p. 48).

Now, according to Llinás, "the brain's control of organized movement gave birth to the generation and nature of the mind" (p. 50). In his view, the intrinsic properties of the brain that swiftly construct and employ muscle synergies for movement execution are in every way similar to those intrinsic properties of the brain that generate the sensorimotor images responsible for perception and cognition (p. 55). There is in the field a large amount of serious research concerning the intimate connection between perception and movement, so this is no news. What is important to stress here is the way this subject is related by Llinás to his idea of the operation of the brain as a *closed system*. The premotor constructs, according to him, must *emulate* external reality in order to determine the consequences of the organism's movements. There occurs some sort of "embedding" of properties of the external world (he calls these properties "universals") into the functional workings of the neuronal circuitry of the brain that generates a predictive image of an event to come that causes the creature to react or behave accordingly (ibid.). What we are facing here is a primarily self-activated system whose organization is geared toward the generation of intrinsic images. But

> Given the nature of the thalamocortical system, sensory input from the external world only *gains significance* via the preexisting functional disposition of the brain at that moment (…) Such self-activating system is capable of emulating reality (…) even in the absence of input from such reality, as occurs in dream states or daydreaming. (p. 57, emphasis added).

If the generation of a "reality" is an intrinsic and internal accomplishment of the brain, what is then the difference between the operation of the brain in sleep and that same operation when one is awake, that is, *conscious*? The difference, in Llinás's words, is "significance". What could "significance" possibly mean in this context?

Sensory occurrences are integrated into whole percepts in a contextual manner, that is, the integration is dependent on the internal context of the brain. During sleep, this internal context "does not grant significance" to the sensory input, which means that it does not exist as a *meaningful* event *for* the system. In wakefulness, to the contrary, the relevant stimuli are "phased together", and in this way become meaningful, so that it makes sense to say that they are given "internal significance". A first point to make at this moment is to notice that "significance" stands here for meaning emerging precisely in this correlation. Now, we may ask: how is "significance" –"meaning"– accorded to sensory events?

No doubt, the description of the mechanism responsible for this constitutes the core of the proposal made by Llinás. There is a sort of "structural a priori", an innate functional connectivity (well-specified neuronal circuits) that is present at birth in all normal brains, composed by the thalamus, the cortex and the specific

connectivity between them. The mechanism that assures that connectivity is not spatial, but *temporal coherence*:

> Building on physical connectivity, the nerve cells of the brain have created an "interlocking" solution: the synchronous binding in the time domain of [those] individual neuronal activities. By making different time-interlocking patterns, neurons can *represent a unity of reality* by combining the individual, fractionalized aspects of reality that each neuron carries. This time-interlocking phenomenon is time coherence. (p. 120–121, emphasis added)

Perceptual unity is believed to be attained by way of this neurological mechanism, known as "cognitive binding", which is very similar to the "motor binding" mentioned above that is believed to be performed by the inferior olive. This mechanism is constituted, according to Llinás, by temporally coherent oscillations at 40 Hz, which have been detected in neuronal activity during cognitive tasks and which are also intrinsic to the thalamocortical system[2]. Conscious experience is then an event "determined by simultaneity of activity in the thalamocortical system" (p. 124).

The Consciousness-World Correlation

I want now to draw some philosophical consequences that seem to follow from this approach of consciousness.

If we assume both that this mechanism gives rise to consciousness and that it is responsible for the "significance" accorded to sensory events, one (although not Llinás) may conclude that *consciousness is essentially a* "meaning-bestowing" *event.*[3]

Let us take a little closer look. We may take a quite straightforward concept of consciousness as referring, as John Searle (Searle 2002) puts it, "to those states of *sentience* or *awareness* that typically begin when we wake from a dreamless sleep and continue through the day until we fall asleep again, die, go into a coma or otherwise become 'unconscious'." What makes the difference between the sleep and wakefulness states is that in the former, sensory impacts coming from outside of the nervous system are not *correlated* with ongoing thalamocortical activity, and thus do not exist as functionally meaningful events *for* the organism (p. 130), while in the latter they do acquire "significance" by being so correlated. Thus, consciousness makes the difference between the two sorts of states by somehow bestowing meaning upon the affecting events occurring in the wake situation, and in this way *enacting itself as a correlation*, that is, a state in which the conscious organism has a "point of view" on those events. Now, what does this meaning consist in, exactly?

---

[2] This mechanism is also put forward by Crick (1994) and by Edelman (2006).

[3] Notice that the "meaning-bestowal" does not depend on the specific details of the electro-physiological mechanism itself, but on its synthesizing, unifying functions (see below).

If we follow Llinás, we should conclude that this "meaning" is made up of, in the first place, sensorimotor *images* of the external world that allow the brain to *plan and guide* the motor behaviour of the organism within his environment in a moment-to-moment *decision making* basis. But we know how far this concept of "internal images" could lead us if it is taken literally. A more subtle formulation in the same vein is that of a useful, internal construct of the external world (p. 120), a "translation" of the features of external reality into internal patterns of activity, serving the same purposes. No matter what formulation is chosen, though, it is always referred to as being a sort of representation of the external world. But notice that there seems to be no additional fact corresponding to the alleged "significance" accorded to sensory inputs by the workings of the thalamocortical system apart from these workings themselves. So, what is it that makes some events "significant", or meaningful, in the wake situation?

The internal "construct" brought up by Llinás, that is, the global activity pattern formed as the outcome of whole modules of neurons electrically oscillating in phase, and whose activities are supposed to "represent" fractionalized aspects of the external world, is meant to guide *action*, that is, intentional and purposeful bodily movements (as opposed to mere physical moving) at the basic level, and conscious behaviour at the higher level. Certainly, if we want to describe this guiding, nothing forces us to say that it is carried out through internal representations of the external world, no matter how literally we interpret this talk about representations. But what remains true is that at this level of our description we are forced to switch our perspective if we intend to account for the kind of fact that is presented to us. What I want to suggest is that at this particular stage we, the describers, begin to be faced with a totally new state of the system, a state in which what is described has a "point of view" (there are things that "matter" to it), and this fact determines that we switch the perspective from which the description has to be made. In effect, the new state of the system is what I would like to call here *the consciousness-world (C-W) correlation*. And with this new state comes the medium of its operation: *meaning (sense)*.

Let me say something about this *correlation*. What becomes apparent in the description of the brain mechanism just described is that its functioning originates the phenomenon of consciousness, as it has been put forward by Llinás, insofar as it determines, specifies and defines what in the ongoing coupling of the organism and his environment is *significant* for him (what "matters" to him); and this is achieved by the *enacting* of what might well be called *his world* (that which makes sense to him). So, there is no consciousness without a correlative significant world, and a world (as opposed to an environment) is always a world *for* a consciousness. In other words, what I want to suggest is that consciousness may well be seen as a state of the functioning of the brain mechanism just described[4]; but it consists essentially in that this functioning sets up what I have called "a point of view", the

---

[4] Or some other that do the same job. In fact, of the entire body, and not of the brain or the nervous system alone, as it will become apparent later on.

point of view of the organism, insofar as it is the organism who determines, specifies and defines, through his very functioning, what "matters" to him, and so at the same time enacts a world *for* him. This correlation, I want to say, constitutes the level of *meaning (sense)*, the level at which every event is not just a physico-chemical happening, but is an event *for the organism* because it "makes sense" to him.

Now, meaning, sense, does not emerge as a finished, pre-configured medium. It has to do essentially with the organism's orientation to movement, which from the perspective of the organism itself is called "action", and is literally *constituted* throughout the various levels at which the embodied organism has to deal with his acting within his current *world*. If the theory that I am considering here is correct, constitution of meaning is partially, though essentially, linked to the achievement of temporal coherence that brings together the intrinsic oscillatory properties of the nervous system in operation into a single, unified and centralized construct, which Llinás significantly calls "a single understanding" (p. 23). Hence, the work of consciousness that makes it essentially *meaning-bestowing* can be seen as one of synthesizing, unifying and centralizing the global activity of the nervous system into coherent, significant and unified *contents*. Notice, these are *meaningful contents* to the extent that the appropriate C-W correlation is enacted and so they enable the organism to *actively* cope with his environment by making a world out of it; and they can do this to the extent that they are meaningful (they have "significance").

Thus, to *design* and *guide* action, as opposed to simply cause physical movement, is, we may say, *the meaningful aspect* of the functioning of a fully embodied brain. I have previously said that the mechanism in the thalamocortical system that gives rise to consciousness is *meaning-bestowing*, because it marks the difference between what has and what does not have "significance" for the organism, in the sense of taking or not taking part in the generation of meaningful action. If the hypothesis considered here is sound, this *meaning-bestowing* feature consists basically in the transformation of sensory impacts into patterns of activity, referred to as *meaningful movements*, or *meaningful motricity*, under the regulations imposed by internal context generated by the activity of the thalamocortical system, especially the non-specific intra-laminar system, and intrinsic movement-oriented brain activity. The *bestowal of meaning* achieved this way and by this embodied mechanism should allow us to call "consciousness" this global electrical oscillation in phase of the thalamocortical system, precisely as far as it is *meaning-bestowing*. That is, as far as it enacts the C-W correlation which I have proposed to call "meaning" ("sense").

I think this "phenomenological" extension of the theory is well suited to overcome some well founded criticisms that have been addressed to it, mainly those coming from the embodied approaches of conscious perception, as they can be found, for instance, in Alva Noë's and Kevin O'Regan's "Sensorimotior Account of Vision and Visual Consciousness" (Noë and O'Regan 2001). I would like to show this by reviewing the main lines of this related account of conscious perception and then to suggest a non substantial amendment of it in order to enhance its significance for the issue that is at stake in this paper, namely the issue of the sense of an external world generated entirely by internal conscious operations.

The Sensorimotor Emergence of the World

There is an extremely interesting account of perceptual experience that is very close to the view I am trying to build on these matters. It is the so called "sensorimotor contingency theory" suggested by Alva Noë and Kevin O'Regan in their celebrated 2001 article. Sensorimotor contingency theory (SMCT from now on) is presented as a theory of visual experience (visual consciousness, actually), but it is immediately extended to apply to the other sense modalities. On this account, vision is not treated as the generation of an internal image of the external world, but instead as an activity, an "exploratory activity" (Noe and O'Regan 1939). The central idea of their approach is that

> Vision is a mode of exploration of the world that is mediated by knowledge of what we call sensorimotor contingencies, on the part of the perceiver. (Ibid.)

In another paper, O'Regan and Myin (O'Regan and Myin 2002) say that SMCT is "a specific skill theory". It is worth reviewing what exactly a "skill theory of experience" is.

The "skill theory" approach to experience, mainly perceptual experience, can be found in different papers by Andy Clark in which he addresses the question of phenomenal feel in connection with sense modalities. According to Clark's "skill theory", as summarized by O'Regan and Myin (O'Regan and Myin 2002), perceptual content is constituted by the behavioral skills available to the perceiver by virtue of perception. So, for example,

> An experience has a specific spatial content to the extent that it allows the perceiver to make appropriate movements with regards to space: to orient, to reach or to grasp. (Clark, cited by (O'Regan and Myin 2002), 29)

The behavior, – the enabled actions – are taken as possibilities. So, "perceptual content is constituted by the *potential* for action enabled by the experience" (Ibid).

I will take this approach as an enhancement and a partial correction of the internalist view of conscious experience exposed here. It enhances and corrects that view by making explicit the main feature of the alleged link existing between perception and action. This link appears to be in fact a constitutive unity made out of experience and possibilities for action. Thus, all the characteristics of conscious experience are literally constituted by the potential for action that the organism possesses. Thus, the internal brain processes described by the internalist account (coherent oscillations in the thalamic cortical system) play an important role in the nervous system activity linked to consciousness only in so far as they contribute to the enactment of the world in the C-W correlation as a world of possibilities for action.

It is well known that the connection between intentional acts, typically perceptual experience, and potential for action has been a main theme in the phenomenological tradition, ever since Husserl himself characterized the core of the intentional relation to the object with the words "Ich kann". What is more interesting for my purposes here, is to remember that what is generated in the basic intentional relation

is not just an experiential content of possible activity, but also, and simultaneously, the world for this activity to be deployed. This "world" corresponds to what Merleau-Ponty called the "niche" created by the subject's deployment of its bodily capacities. Myin and O'Regan mention this idea of Merleau-Ponty's, but I think they do not fully exploit it in the sense I expect to show in a moment. It also seems to me clear that this "niche" corresponds to what may be called "the cognitive domain" of the subject.

The SMCT proposed by Noë and O'Regan is presented by the latter and Myin as a specific skill theory. It was proposed as a means to deal with some major issues existing in theories of consciousness, especially the question of the origin of conscious experience as well as that of how to provide an adequate account of some of its features, particularly phenomenality and differences between modalities. I am not interested now in these issues, so I will not touch them. I just mean to extract some important aspects of the theory that can be useful to deal with the problem of the subjective constitution of the external, objective world.

> Sensorimotor contingencies (dependencies, co-variation) are the laws that link a perceiver's actions to the changes in sensory input that these actions cause ((O'Regan and Myin 2002), 33).

Consider the example of seeing a line: the projection in the eyeball remains identical when you move your eye along it, but changes drastically when you move your eye in any other direction. This law of co-variation is a sensorimotor contingency applicable to any seeing of a line.

According to this,

> Perceiving … is an organism's exploration of the environment that is mediated by knowledge of sensorimotor contingencies (O'Regan and Myin 2002, 34).

Perception is thus essentially linked to the potential navigational movements of the perceiver. The link is essential in that

> With any exploratory movement that the perceiver makes, she has knowledge about how input will change.

This knowledge is supposed to be the knowledge of a *law*. The authors are aware of the implications of calling it "knowledge", and so they make the claim that it is *implicit* knowledge. This means that it "is present only in the particular ways the ongoing exploration unfolds" (Ibid.) For me, the following formulation is even more clear:

> Knowledge of particular sensorimotor contingencies should be understood as literally constituting a perceptual skill, analogous to a skill such as tying one's shoelaces.

Most of the time we behave in an accurate way in our ever-changing environment. Our spontaneous actions are most of the time coherent with the demands of the situation, as if we always knew what to do. This fact can be explained by the continuous coordinating of one's actions with opportunities to act offered by the environment. Now, to be guided by the sensorimotor contingencies does not amount to being fully perceptually aware. To get perceptual awareness, the perceiver must deploy

A further skill, namely the skill of integrating one's purely perceptual skills into one's intentional behavior (35).

This means that in any perceptual situation the perceiver becomes perceptually aware of an object of perception

when she focuses on one of the ongoing patterns of exploration and thus have intentional access to certain currently applicable sensorimotor contingencies. (Ibid.)

The SMCT is for me a very appealing approach to conscious experience. So, the relatively small amendment I am going to propose now is not intended to make corrections, or to suggest any major change on it. I will just try to use it in a profitable way in order to face the problem of the subjective constitution of the external, objective world, which is my problem in this paper.

The Shared World

How is it that we have the *sense* of an external reality? I mean: not that we **believe** that there exists a reality outside of my mind; that can be contended, as the history of philosophy can easily show. What cannot be even discussed is that we live with the sense, the pre-noetic structure of experience of living in a world that is independent of our mental life. The question is: how is it?

Skill theories of perception give us some important clues to try to answer this question. Epistemological anti-realism, i.e., skepticism, can enjoy some plausibility mainly because of the model of perception that has been dominant in the epistemological disputes. This model corresponds roughly to what has been called the "cognitivist" view of perception and cognition. According to this view, perception is supposed to build internal representations that in some sense reproduce an external pre-existent configuration of the world, via the senses and internal activities of the nervous system. But it is not an incoherent view to suppose that these internal images generated by the activity of the nervous system do not actually correspond to anything outside the mind. So, how can we be certain that that pre-configured external world really exists?

Skill theories, and the SMCT in particular, give us a different picture. For these theories perception is a kind of *doing* in the first place, i.e., acting in an environment. Thus, to be perceptually aware of something is not to entertain an image of that thing, but to interact with it, and in this sense the question of how we come to know that there is an external, independently existent world should not even arise. Because it is a fact that we do not *know* it. We simply act.

Now, from a phenomenological perspective, this account could possibly be qualified as "insufficient", in the sense of not being the whole story. What is missing here, in my opinion, is a correlative account of the constitution of the intentional correlate of perceptual, conscious experience, that is, the perceived, experienced, lived world: the "niche". Specifically, an account of how this world becomes constituted as an external, independent world in the very same process of experiencing it.

This correlation is a phenomenological datum and I think any valid theory of conscious experience should account for it.

I would like to outline now a possible path to be followed in order to give an answer to that question. It is a kind of speculation, because I do not have many arguments to support it, except that if it proves to be well oriented, some pieces of the puzzle would probably find their places.

I think that what gives us the originary, pre-noetic sense of an external reality (the world) is something that should arise from our early experience of others. And I think also that the main spontaneous strategy we follow to achieve this is something very similar to what Donald Davidson has called "triangulation". I mean by this that the pre-noetic sense of an external, independent world is constituted, not in first-person performances of a nervous apparatus, but rather in the continuous triadic interaction between self, the other and something happening in their immediate environment and being shared by them. Thus, the lived world that is always present as the correlate of our conscious life is not actually the outcome of a subjective, first-person constitution, but rather it is continuously enacted in an inter-subjective, second-person constitution. That means that the world co-originates with conscious experience only as a second-person experienced world.

This original sense of externality becomes constitutive of our very sense of the world, because every experience occurs against this very background and because any further experience does nothing but confirm it. Worldly experience in humans comes with a sense of being a shared experience. It is this sharing feature that we should find, I speculate, at the origin of the attribution of mental states to others, but also to oneself. I think there is sufficient empirical evidence supporting this claim in studies on developmental psychology and primatology.

I have called "triangulation" the main strategy I want to postulate to account for the second-person constitution of the external world, and I even mentioned the Davidsonian lineage of the term. Nonetheless, I do not want to suggest, not even implicitly, that "triangulation", in the sense I am now using this term, is some kind of conceptual device. It isn't, obviously. And the main reason why it is not is because the early experience of others in which triangulation takes place is perception of the other, and we have already seen that perceiving (and in this particular case it involves almost every sense modality) is always a doing something. This fact suggests to me that this early experience should be conceived of as a sort of *coordination of actions*. Something like *a dance*.

In effect, one can speculate that in the very early experiences of others, "triangulation" spontaneously emerges as a matching of first- and third-person perspectives in the constitution of a coordinated *activity*. I am thinking in of a situation in which an adult and an infant are engaged in a shared perceptual activity in which the infant makes the experience of the first-person of self and the third-person of other. If the skill theories of perception, and particularly the SMTC are right, one can imagine that this latter perceptual experience swiftly becomes a mutual coordinating of potential exploratory movements of both subjects. But the shared perceptual activity, insofar as it is directed at (referred to) some third object, must be too a mutual coordinating of potential movements, with the difference that

in this case there are sensorimotor contingencies that are linked for both subjects to a third object. For this object to become a shared perceived object both subjects need to match their first- and third-person perspectives and this is what triangulation is supposed to perform.

I conclude with this single remark: current embodied experience of the everyday world involves the development of an intersubjective, social cognitive achievement. This consists essentially in actively coping with the world by way of sharing and coordinating actions with others. In this way, the world for us has the intrinsic sense of being a shared world, that is, a world that is not the content of my cognitive, mental states, but just the medium of our intersubjective, coordinated activities.

References

Crick F (1994) The astonishing hypothesis. Simon & Schuster, New York

Edelman G (2006) Second nature. Yale University Press, New Haven, CT

Llinás R (2001) I of the vortex. MIT Press, Cambridge, MA

Noë A, O'Regan K (2001) Sensorimotior account of vision and visual consciousness. Brain Behav Sci 24(5):939–973

O'Regan K, Myin E (2002) Perceptual conciousners, Access to Modality and Skill Thories. J Conscious Stud 9(1):33

Searle JR (2002) How to study consciousness scientifically. In Consciousness and Language. Cambridge University Press, Cambridge

Part VI
Language and Meaning

Husserl and Language

Peer F. Bundgaard

Preamble: Language and Husserl

From a purely quantitative point of view, Edmund Husserl has devoted a rather small amount of time and space to the study of language proper. Essentially, his contributions within this domain amount to the description of *language use* in the First *Logical Investigation* (Husserl 1901), and the determination of the essential properties of *language as such* (independent of any specific use) in the Fourth *Logical Investigation*. Otherwise, language is only sparsely dealt with in Husserl's writings: the unpublished note "On the Logic of Signs (Semiotics)" (Husserl 1890) anticipates the distinction between "expression" and "index" which constitutes the starting point of the First *Logical Investigation*; i.e. the difference between a linguistic or any other sign bestowed with intentional meaning and any type of sign which is immediately or physically linked to its meaning: smoke → fire; scar → wound; weathercock → wind, etc.); *Formal and Transcendental Logic* (Husserl 1929) contains an appendix related to the theory of syntax outlined in the Fourth *Logical Investigation*; and, finally, a number of passages from *Experience and Judgment* (Husserl 1939) reexamine the relation between perceptually formed or antepredicatively structured meaning and its linguistic, predicative articulation (as we shall see this issue is also in the heart of the discussions unfolded in the Fourth *Logical Investigation*).

Husserl's theory of language and his analysis of linguistic meaning seem, thus, to have lived a rather insular life in the universe of his thoughts and writings. Actually not only there: few philosophical scholars have paid systematic attention to and emphasized the importance of Husserl's investigations into the nature of language and intentional language use[1]; and even fewer have pinpointed its import

P.F. Bundgaard
University of Aarhus, Center for Semiotic Research, Denmark

[1] Remarkable exceptions are Benoist (e.g., 1999, 2001, 2002), Gardiès (1975), Mulligan (e.g., 1988) and Smith (e.g., 1987).

on linguistics proper. There are a number of reasons for this. The characterization of language undertaken in the Fourth *Logical Investigation* – "The Distinction between Independent and Dependent Meaning and the Idea of Pure Grammar" – is truly ambitious in both depth and breadth. It aims at establishing the essential properties of language as such, independent of any user and independent of any given instantiation of the species 'language.' It is developed explicitly in the vein of *la grammaire raisonnée* of Port-Royal and is in many important respects a forerunner for both Roman Jakobson's structural linguistics and formal, universal grammars of a Chomskyan sort.[2] Nevertheless, the extension of its ambition notwithstanding, it has traditionally been considered or reduced to being a straightforward application of the part-whole analysis and the conceptual tools and principles gained and laid down in the Third *Investigation*. In this respect, a sort of illustrative flavor seems to cling to it: it shows how the mereological a priori from the previous investigation can be shown to govern one among other possible domains: in this case, language. Secondly, Husserl never really readdressed the issue in his future work, at least not systematically and in terms of a pure universal morphology for language as such, a circumstance which, again, enhances the somewhat confined and, as it were, marginalized character of Husserl's investigations into the nature of language. The essence of language seems to have been established in passing, incidentally.

Structure in Language and Function of Language

The study of language as such – i.e. irrespectively of its concrete instantiations as English in eighteenth century, Finnish yesterday, Swahili tomorrow, etc. – can follow two directions. Firstly, it can be characterized in terms of the properties which make it a self-contained symbolic system: i.e. the rules that govern the correct formation of sounds and sound patterns and the correct formation and combination of word signs. In this perspective, meaning is an upshot of linguistic well-formedness – when words combine right, they mean all-right – and the (arduous) task consists, then, in defining the principles and rules of well-formedness. The second direction along which language can be defined is with respect to the use people make of it in order to express themselves: language is a symbolic system and as such a tool for communication; thus, words take on a meaning according to speaker's meaning intention and conforming with the specific use speakers make of them in order to

---

[2] It is, thus a counterpart to Saussure's synchronic determination of *langue* as the proper object of the new science he meant to inaugurate. However, in his writings, Roman Jakobson suggests that it has had much more immediate bearings on the development of structural linguistics than Saussure's *Cours de linguistique générale* (Saussure 1915) since it provided with a principled, positive determination of "structure" in terms of dependence relations. Moreover, the essence of language as such in Husserl is not *semiological*, its cornerstone is not the sign as it is in Saussure; rather, the heart of language as such is *syntactic*: its primitive constituent is the syntactic construction (cf. section on The Syntactic A Priori below).

express what they have in mind in a given situation. Along this dimension the essential properties of language are not inherent in language, or reducible to the language system, they transcend language: they are defined by those links that connect language and language use to those mental acts whose quality, content, and object reference they express.

A remarkable character of this divide – which, basically, corresponds to the semantics-pragmatics divide in linguistics and philosophy of language – is that the analysis of language carried out in either one of these domains does not seem to have any consequences for or to be constrained by the analysis fulfilled in the other domain. If our concern is to define those properties of language construed as a formal symbolic system which rule the combination of partial significations of different sorts into organized meaning wholes, and if we do so by considering language as a higher order abstract entity, then, of course, the use empirical individuals make of it, what they mean by their words and sentences, and how they come to make sense of them, all this is, if not irrelevant, at least only of subordinate importance: of course people use words and sentences in this and that sense, but it is still a fact that sentences can be used only if they have been constructed in a principled way. Therefore the crucial question remains: what are the principles ruling the combination of sounds into words and words into sentences? On the other hand, if a priori, by essence, language is a tool for communication, its essential properties can never be reduced to its grammatical, immanent properties: grammar does not tell us how people intend to mean this and that by using these and these words with this and that meaning in this and that context: grammar alone does not reveal the crucial characteristic of language, namely that it takes on specific meaning only with respect to speaker's intention. Therefore, along this line of thought, the decisive question reads: how does language (and grammar) hook up with mental acts and their contents?

In short, we have not only two types of definitions of language, but two essentially different types: what seems essential along one direction of determination seems accidental along the other; what language is immanently does not seem to have any import on the determination of the use intentional subjects make of it: the design features of language as such – the grammatical a priori – and the design features of language use – the ontology of language use – seem to lie each in their own dimension.

For decades, if not simply from the beginning of the last century, this divide has run through linguistics and philosophy of language as a foundational schism, or more trivially as a theoretical iron wall. It manifests itself in the historical conflict between formal grammars (mainly Chomskyan and post-Chomskyan grammars, but in a continental perspective also linguistic structuralism) and functional/pragmatic or cognitive grammars, and is at the heart of Merleau-Ponty's distinction between *parole parlée* (speech as formed meaning) and *parole parlante* (speech as active meaning formation). As is often the case for such absolute divides, this one also has historically invited people to choose their camp in an either-or way, thus hindering the cross-over investigations into the internal links between immanent properties of the language system and the possibility of expressing mental acts by means of language.

Edmund Husserl's theory of language is no exception to this state of affairs. In fact, it even seems to have set up the framework for it. Husserl, in his First and Fourth *Investigations* carefully displays the tenets of both an ontology of language use[3] and the principles of the grammatical a priori. In the First *Logical Investigation,* linguistic meaning is mainly defined within a communicational context, with particular focus on speaker's meaning intention. Here, thus, the general components and levels of language use are laid down. In contrast to this, the Fourth *Logical Investigation* defines lingustic signification as an upshot of grammaticality: words mean something whenever (or if and only if) they appear in sentences with the right syntactic format. In the First *Logical Investigation* meaning is defined in terms of the semantics *of the intentional act* and the task is one of accounting for the acts that assign meaning to and grasp meaning from linguistic expressions; in the Fourth *Logical Investigation*, however, meaning is defined in terms of the *syntactic* rules governing the sequencing of words into the right sentence format, regardless of their actual meaning and regardless, thus, of speaker's meaning intention.

These developments are not in themselves antagonistic, at least not in principle. Even though different, antagonistic types of linguistic theories and different types of philosophies of language have developed on the grounds of the difference between these two aspects of the object, it remains nevertheless a fact that language is governed by syntactic rules (principles of composition), and that language meaning cannot be correctly characterized without taking into consideration speaker's meaning intentions and the semantics of his mental acts. However, Husserl seems to have left unanswered a question in dire need of clarification: if language can be determined along two equally essential lines of characterization, isn't it then of crucial importance to account for the compatibility of the latter? If we have one form of principle governing meaning articulation in one domain – say, the immanent combinatorial properties of language as a self-contained formal system – and another principle governing meaning articulation in another domain – say, the general features and levels of language use – isn't it then our task to explain how meaning structured in one modality can be articulated in the other? In short, how can language, by virtue of its immanent structural properties, express what has been structured and organized in the mental acts which underpin language use? What makes language capable of expressing, in a linguistic format, an intentionally formed meaning?

Although few commentators have observed this, Husserl has in fact outlined an answer – a both bold and all too vague answer – to this question. In the following sections we shall therefore in a rather flatfooted way follow Husserl's determination of language along the, if not divergent, then at least apparently parallel lines described in the First and the Fourth *Logical Investigations*. Language defined within a mainly communicative context; language defined as an autonomous formal symbolic system. Then, we shall see how Husserl in a rather overseen section of the Fourth *Logical Investigation* (in fact the first 9–10 paragraphs of the investigation)

---

[3]The expression is from Smith (1987).

lays bare a level of semantic structuring in language whose governing principle could account for the faithful articulation of pre-linguistically organized meaning and the linguistic format within which it is expressed.[4]

It is, in a sense, a delicate task to assign particular importance to these first 9–10 paragraphs of Husserl's investigation and a *caveat* should therefore be made: Husserl does certainly not, either explicitly or tacitly, introduce his ideas in these sections as constituents of a theory of language which is essentially different from the one carried out in the subsequent paragraphs; on the contrary, any and all bits and pieces of his analysis are presented as leading to the same, unitary "pure logical grammar." It is nevertheless relatively easy to show (1) that the first part of the analysis is semantic, the following syntactic; (2) that the two accounts ride on essentially different types of principles ruling the composition of partial significations into organized semantic wholes; and (3) that the Fourth *Logical Investigation* is incongruous in this respect. Yet, the internal incongruity of the investigation did not prevent Husserl from making, in our eyes, a crucial contribution to the theory of language, one which makes it possible to correlate the design features of language use and the design features of language as a grammatical system, i.e. one which combines the structure and contents of the mental acts with the structure and semantic forms of the symbolic system which express the former.

First *Logical Investigation*: the Ontology of Language Use

This section introduces to the constituents of an ontology of language use, namely (1) intimation; (2) the meaning conferring acts; and (3) the meaning fulfilling acts.

Intimation

In the very beginning of the First *Logical Investigation*, Husserl draws a seemingly scholarly distinction between 'indications' (*Anzeichen*) and 'expressions' (*Ausdrücke*): the disentanglement of the two sign types is nevertheless important because, even though both trigger an interpretive representation, only the latter express a subjective meaning intention. In the communicative context the distinctive feature of a word sign or an expression is therefore that it is used by a speaker "with the intention of 'expressing himself about something' through its means: he must endow it with a sense in certain acts of mind, a sense he desires to share with his auditors" (First Logical Investigation §7, p. 189). For the hearer, the difference

---

[4]To my knowledge, only Jocelyn Benoist (1999, 2001, 2002) has remarked that Husserl in fact unfolds *two* different grammatical a priori in the Fourth *Logical Investigation*, one syntactic – as it is generally acknowledged – and one semantic-mereological. Benoist, however and contrary to what will be claimed here, dismisses the latter as inconsistent (cf. Bundgaard 2004a, b).

between an indication and an expression is that the latter does not only mean something, but is also used by someone to mean something in particular. Hence the first conclusion: the zero-level of linguistic articulation of meaning is pre-linguistic; in order for a word sign to take on a semiotic function, the hearer must recognize that the utterer intends to express himself through it: before we capture the meaning of the sign, we capture the fact that the sign expresses a meaning intention. Notice that this constraint is, by and large, independent of any criteria of well-formedness: this is, trivially, the reason why, for example, infants with still rather incomplete phonatory skills may fail to comply with the rules of pronunciation and still be considered as expressing themselves by means of word signs, whereas parrots, however skilled in recitation, will never surpass the threshold of shear sound making. In a nutshell, from the perspective of language use, words may comply with all possible phonological, semantic and syntactic rules, they do nevertheless not mean anything if they are not used and interpreted as announcing a meaning intention. This fundamental function of word signs or verbal expressions is what Husserl calls their "*intimating*" function: it announces speaker's intention to express himself, and it does so at a level that precedes the grasping of any actual, linguistically formed meaning:

> What first makes mental commerce possible, and turns connected speech into discourse, lies in the correlation among the corresponding physical and mental experiences of communicating persons which is effected by the physical side of speech. (First Logical Investigation §7, 189)

As suggested, it is in fact the other way around: it is only insofar sound patterns display an intimating function, non-linguistically, by virtue of their manner of givennes, that they are taken to mean something *stricto sensu*:

> [t]o understand an intimation is not to have conceptual knowledge of it, not to judge in the sense of asserting anything about it: it consists simply in the fact that the hearer intuitively [anschaulich] takes the speaker to be a person who is expressing this or that, or as we certainly can say, perceives him as such. (First *Logical Investigation* §7, 189)

Notice that language is here, from the outset, a priori, embedded in an overarching intentional framework. The fact that words and sentences pronounced by parrots fail to be considered as expressions – even though they are acknowledged as tokens of words and well-formed sentences – tends to show that linguistic signification is overdetermined by intentional meaning: according to Husserl, it takes a meaning intention to get an actual word meaning.

The Meaning Conferring Acts

Obviously, from the speaker's point of view it is an analytical distinction to differentiate acts which express an intention to communicate by means of word signs from acts which use the same words signs to refer to this and that state of affairs through this and that specific meaning. In other words, the wish to communicate (intimation) implies the wish to communicate something in particular (by means of

meaning conferring acts). When confronted with a sentence like '*He is red*', the hearer may grasp the intimating function while still doubting about the concrete meaning of the sentence, whereas the speaker of course knows what meaning he has conferred to it (say, "He is a Native American/a communist/exhausted/ashamed"). In this respect to intimate a communicational intention is simply to give or confer meaning to meaning signs which then are used to represent a state of affair in a specific respect.

The relevance and importance of this distinction, and thereby of the meaning conferring acts, can be made clear in two ways. Firstly with respect to receiver's point of view. Often, as just suggested, we do grasp a clear communicational intention without grasping the matter of its purpose, or we doubt whether we got it right, or we simply misinterpret it. In all such cases, the grasping of the intimating function is constant and correct, but the correct interpretation of the communicated matter does not obtain: we do not get the meaning conferring acts right, we cannot see how the corresponding meaning obtains. So to grasp the meaning of an utterance is to represent the meaning conferring acts.

The second reason why meaning conferring acts are essential as such is important from a both linguistic and cognitive point of view; it further emphasizes the fact that all linguistic commerce is intentionally framed; the point rests on a simple observation: to confer meaning to word signs is not simply to overtake their signification as it is laid down in a dictionary, nor is it to assume their pre-established truth value in a communicational context. To confer meaning to word signs and sentences is a much stronger and much more literal affair, since word signs are by essence underdetermined, both *intensionally*, *intentionally*, and *extensionally*, i.e. with respect to their possible signification, with respect to the way they are used to qualify an object, and with respect to inherent object-reference (truth value). Semantically or intensionally, no word sign can be assigned one and only one meaning, they always display a network of more or less interrelated meaning nuances; intentionally, no word sign and no combination of word signs, however simple, specify in and by themselves a univocal relation to any kind of object; logically, no word can be said to have any reference value other than a default one. In short, words and sentences have only default significations and are likely to take on different meanings in different situations (refer to different kinds of objects): by virtue of speaker's meaning conferring acts.

Word signs are not only semantically underdetermined in more or less exotic cases of words or more less exotic uses of language; say, different, but homonymous words as '*bank*' (what is green in '*the bank over there is green*'?); or meta-uses laid bare in the difference between "This" *is an example of a demonstrative pronoun* vs. '*This is an example of a demonstrative pronoun*.' Cases like these do indeed also nicely illustrate the fact that the signification of words and sentences is not univocally encapsulated in them and, thus, that their relation to the object does not obtain automatically; but, trivially, the same can also be shown to hold for very plain, perfectly non-exotic expressions such as "The ball is green", which may, in different uses, mean (a) its surface color is green; (b) (said by a referee in a basketball game) it should be played by the green team; or "This computer is child safe",

meaning (a) children can safely use it to surf on the internet; (b) it is solid enough to resist children's careless use.[5]

Husserl does not comment (or hardly does) on the meaning-conferring acts with explicit reference to the intensional, intentional, and extensional underdetermination of word signs and language. It is nevertheless a simple correlate to the claim that expressions take on an actual signification only insofar as meaning is intentionally conferred to them, i.e. only insofar as they are used to express the way this and that object or state of affairs has been intended. In the light of these developments we can now propose a finer grained definition of the distinction between the intimating function of the word sign and the meaning conferring acts by virtue of which it takes on a specific meaning: given one and the same expression or combination of expressions,

- The intimating function is what establishes, in the expression, a relation to the utterer (utterer's intention to communicate).
- The meaning conferring acts, in turn, establish the relation of the expression to the object (the nature of this relation, the way in which the expression is meant to refer to the object, is then the intended meaning of the expression).

The latter relation between meaning and object such as it is laid bare in the description of the meaning conferring acts is thoroughly examined in the First *Logical Investigation.* Here "relation" essentially entails that meaning and object are not the same ("the object never coincides with the meaning", First *Logical Investigation,* § 12, 197). This does not imply that meaning and object are unarticulated, on the contrary we do have access to things in the objective world, but always through the specific way we intend them:

> […] an expression only refers to an objective correlate *because* it means something, it can be rightly said to signify or name the object *through* its meaning. An act of meaning is the determinate manner in which we refer to our object of the moment, though this mode of significant reference and the meaning itself can change while the objective reference remains fixed. (First *Logical Investigation* §13, 198)

As mentioned before, this mode of attending to things, this manner of intention and content of intentional experience is the very meaning we confer to word signs when we use them to name objects and express our experience of objects. We can now propose a simple definition of what Husserl would consider the meaning of an expression in a communicational context, that is to say the meaning conferred to a word by someone in order to refer to some object: the meaning of a combination of word signs[6] is the (pre-linguistic) "matter" of the corresponding intentional

---

[5]Examples like the above abound in Cognitive Linguistics (cf. section on Language and Conceptual Structure – Evidence from Cognitive Linguistics below). Even though they sometimes appear to be invoked in order to challenge the referential function of language *tout court*, this sort of infelicitous conceptualism does not follow; a Husserlian intentionalism is all as possible: we use plurivocal words to express univocal thoughts about things.

[6]It follows, then, that the meaning of each words sign is relative to the part of the global representation (the whole "matter") it is intended to evoke.

act, the thing *as* it has been intended (in the sense of Husserl's act quality-act matter analysis from the Fifth *Logical Investigation*). This, of course, explains why one and the same expression or combination of expressions can refer to different objects (namely, through different meanings); or why expressions can refer to exactly the same object, yet with different meanings (exactly as in intentional experience we can have acts of the same quality directed toward the same object, but with different act matters, i.e. they intend the object through different contents (by giving them different "aspectual shapes" as John Searle would say)). Examples of the latter are legion and post-Fregean standards: 'the equilateral triangle' vs. 'the equiangular triangle,' 'the victor at Jena' vs. 'the vanquished at Waterloo' – significantly, the latter two expressions are examples that Husserl not only uses, here, in the First *Logical Investigation* in order to pinpoint the difference between meaning and object, but also in the Fifth *Logical Investigation* to illustrate the relation between act matter and object. Or in Husserl's words: sentences like the above

> are not merely grammatically, but also 'cogitatively' [*gedanklich*] different, i.e. different in meaning-content. But they express the same state of affairs [*Sachlage*], the same 'matter' (*Sache*) is predicatively apprehended and asserted in two different ways. (First *Logical Investigation*, §12, 198/48)[7]

However, it is possible to give more straightforward, less Fregean examples of sentences with the same reference value and different meanings. Consider, for example, 'a *is bigger than* b' vs. 'b *is smaller than* a'; or 'a *is parallel to* b' vs. 'b *is parallel to* a'. Or finally 'a *receives x from* b' vs. 'b *gives x to* a.' The latter examples are different from the "Fregean" cases both epistemically and ontologically. Epistemically, each expression necessarily entails the other ('*a* is parallel to *b*' of course implies the inverse), whereas no knowledge about Bonaparte being the vanquished from Waterloo entails insight in his being the victor at any other place. Ontologically, no property of the object or the referent scene referred to in any one of the sentences is not also referred to in the other: the difference in meaning, laid bare by these examples, is thus clearly a difference pertaining to the way in which attention has been directed to the object by the language using subject: each of the expressions specify where speaker has placed his attention (on the element expressed by the grammatical subject). In short, word signs to which meaning has been conferred not only designate objects or aspects of objects, but also aspects of the mental acts in and through which the objects have been experienced, i.e. speaker's mode of attention, manner of directedness toward the object.

---

[7]The German terms have been added because the English translation lends itself to misinterpretation: *Sachlage* (or 'situation') seems here to be used in much the same sense as in *Experience and Judgment* where it is the objective correlate of any perceptual judgment giving rise to the constitution of a meaning-content (or a *Sachverhalt*, state of affairs). Thus '*a* is parallel to *b*' and '*b* is parallel to *a*' express different states of affairs, but name the same *Sachlage*. Similarly 'matter' in the English translation should not, on the contrary, be confused with 'matter' as it is used in the Fifth *Logical Investigation*.

So far language use has been seen to be "bound up with mental acts", as Barry Smith has put it (Smith 1987). This implies, of course, that the question about what language is (what rules govern the formation and combination of word signs) is overdetermined by the question about how linguistically articulated meaning obtains (by virtue of what fundamental setup involving mental acts, meaning, and object). In a nutshell (and notice that this is by no means self-evident – Husserl will indeed propose an entirely different definition only three investigations farther on): any property of the language system in charge of articulating meaning only comes to express meaning within an extra-linguistic framework, involving mental acts with specific qualities and matters.

Before confonting the by now urgent question: "How can language faithfully express mental acts and their contents?" we shall briefly mention the last constituent element of language use in a communicational framework: the meaning fulfilling acts.

The Meaning-Fulfilling Acts

Intimation and meaning-conferring acts are "essential" to the word sign (First *Logical Investigation,* §9, 192) in that all it takes for suitable formed sound patterns or written marks to be expressions are acts which use these expressions to name a relation to an intended object: the meaning-conferring acts. However, if we consider the phenomenology of language use, i.e. the comprehensive context of communication, meaning construction also entails a receiver's, or a possible receiver's, understanding of the expressions, i.e. the expression's relation to the object it names through its meaning. Those acts, in the receiver, which realize, fulfill or, plainly, represent the relation of the expression to the object, are, in Husserl's terms, the meaning-fulfilling acts. Such acts should be stratified however, and probably characterized in more detail than Husserl himself does. At least the following cases obtain. The bedrock of meaning-fulfilling acts is plain perception: in some cases the relation to the object is immediately confirmed, in the communicational context, through perception, as in: "See that elderly looking man over there with the green cap?" In other and indeed most cases, the meaning-fulfilling acts exploit the memory system (they are recollections, memories or stored knowledge); to understand speaker's expression, or to realize the expressed relation to the object is to produce a suitable, but not immediately perceptually confirmed representation which may match the intended relation. The range of meaning-fulfilling acts that confirm the expressed relation to the object on not immediately or simply on non-perceptual grounds is vast: it covers acts of understanding recruiting everything from the most immediate recollection to the most formal knowledge. Its limit case is – of course only from receiver's point of view – the expression of a pure, empty meaning intention, i.e. whatever sufficiently well-formed expression which, in lack of any further specification, can take on only a default signification: the elderly looking man with the green cap mentioned

above is an example of a complex expression which in the present context can only be matched with a general, vague representation (implying some "manness," some "elderliness," some "capness," and some "greenness").[8]

Now, if by "ontology of language use", we mean the essential elements involved in the constitution of a sign, and by "phenomenology of language use" understand the essential elements involved in the expression of meaning in a communicational context, then, for sure, meaning-fulfilling acts are not indispensable from an ontological point of view, but they are so from a phenomenological point of view. When a sign is used to express a meaning (a relation to an object) to someone else, the purpose is to share this meaning with the receiver, that is to say to evoke in the receiver a representation that matches the meaning conferred to the sign or combination of signs. In all felicitous – which means in all non-problematic cases of communication – meaning-conferring and meaning-fulfilling act are therefore fused. Husserl resumes it like this:

> We shall, on the one hand, have acts essential to the expression if it is to be an expression at all, i.e. a verbal sound infused with sense. These acts we shall call the *meaning-conferring acts* or the *meaning-intentions*. But we shall, on the other hand, have acts, not essential to the expression as such, which stand to it in the logically basic relation of *fulfilling* (confirming, illustrating) it more or less adequately, and so actualizing its relation to the object. These acts, which become fused with the meaning-conferring acts in the unity of knowledge or fulfillment, we call the *meaning-fulfilling* acts. (Husserl, First *Logical Investigation,* § 9, 192)

Moreover – and still within the phenomenology of language use – the determination of the import of the meaning-fulfilling acts can be further sharpened: insofar as the expression intends to trigger a representation that matches the meaning it expresses, Husserl simply says that "the *fulfilling act* appears as *the act expressed by the complete expression*" (*ibid.*): the meaning of any given expression would then be the representation that confirms or realizes or illustrates or represents its relation to the object.

Partial Conclusion

In this approach to language the exact make-up of the symbolic system used to convey the intended meaning is by and large ignored: every essential constituent of

---

[8]The fact that language makes default sense "in the void," regardless of context, points to the existence of a pure grammatical *a priori*, which is defined independently of any intentional framework and, thus any mental acts in the above sense: it is the *a priori* that makes *possible* (but unspecified) meaning follow from syntactic well-formedness. The *grammatical a priori* also has a cognitive correlate, namely the level of pure *grammatical intuition*: the capacity of recognizing physical manifestations (sound patterns or marks) as tokens of possible language use, since formed according to the rules. It is thanks to their grammatical intuition that people find parrots funny even though their acoustic blasts are devoid of meaning proper. The grammatical *a priori*, in this sense, is what the Fourth *Logical Investigation* sets out to establish.

the ontology of language use is outside language; in mental acts of different sorts, in intended acts, in ways of making signs refer to objects in this and that respect, with or through this and that meaning, inviting receiver to attend to objects with respect to or through that very same meaning, in accordance with that very same aspectual shape, etc. In a nutshell: we have intimation, meaning-conferring acts, meaning-fulfilling acts, but no purely symbolic structure, or hardly, namely nothing but the baseline condition that signs should be combined in a principled way in order to *possibly* mean something.

Now two issues – at least – are left open.

1. The *grammatical* a priori. We know more about language use (its ontology and its phenomenology), but still ignore almost everything about language itself, independently of its use, as a higher order object in its own right: what qualifies it as a symbolic system, what principles allow language in general to express meaning in general (in the void, in a default way, with or without meaning-fulfillment)?

2. *The relation between ontology of language use and grammatical* a priori. If word signs are underdetermined, which implies that they only express a meaning at large, as it were, in specie (if "in specie" means a more or less structured horizon of default significations), then it seems urgent to explain how, on grounds of which cues, do receivers hit the right understanding of an utterance, and do so with such ease? The case is indeed that meaning-intention is specific, particular, the corresponding meaning-fulfillment is specific, particular – yet, the link between the two consists of expressions which in and by themselves do not specify any singular signification nor any specific relation to the object. How do we get from one particular meaning (intention) to another particular meaning (fulfillment) through an open set of possible significations? The stakes are the following: either we can identify a level in language where meaning intentions can be encoded – where the intentional form of apprehension can be linguistically specified – and thus establish an at least partial match between intentionality and semantics; or linguistic semantics degenerate to contextually determined language games; in the latter case, the only essential characterization of language as such would be the one that captures and defines the principles that govern the composition of meaning units into well-formed syntactic wholes.

Both these issues are addressed by Husserl in the Fourth *Logical Investigations*: the former up-front (the purpose is to establish a pure logical grammar, the "ideal scaffolding" of language, as he puts it), the latter explicitly, since Husserl is overtly concerned by determining the properties of language which allow it to faithfully represent our ideas. As already mentioned the exegetical problem consists in the fact that the issues are dealt with along two different directions, one syntactic, another semantic, and it seems, thus, rather daring to consider Husserl's investigation as congruous in this respect. Now, however yawning the gap between syntax and semantics, between the grammatical a priori and the ontology of language use may remain in Husserl's Fourth *Logical Investigation*, it seems clear, in the rearview mirror of recent and not so recent linguistic research, that his outline of

a semantic-mereological a priori in the first 9–10 paragraphs of the investigation makes out an original first attempt to show, on semantic grounds, how language structure is bound up with mental acts and their contents, and, inversely, how the latter can and must be hooked onto grammatical structure and thus be faithfully re-articulated.

The Semantic-Syntactical Duality of the Fourth *Logical Investigation*

The Fourth *Logical Investigation*, "The Distinction between Independent and Dependent Meaning and the Idea of Pure Grammar", is first and foremost known for, if not establishing then at least asserting the existence of a purely autonomous and essentially *syntactic* grammatical a priori. It aims to lay down the laws governing the internal unity and formal congruity of linguistic parts *qua* tokens of syntactic categories; that is to say, independently of what language is used to symbolize or mean, and independently of the embodied human beings that use this symbolic vehicle to express their experiences. The claim is, as is easily seen, developed entirely independently of the intentional framework for language use laid down in the First *Logical Investigation,* and is, indeed – as many scholars have observed – the theoretical source itself of Chomskyanism in both linguistics and cognitive science.[9]

---

[9]One of the first philosophers to consider the Fourth *Logical Investigation* as a precursor of modern logic – and incidentally of modern linguistics – was Bar-Hillel, who in his paper from 1957, "Husserl's Conception of a purely Logical Grammar," concludes: "[…], we may say that Husserl's conception of a purely logical grammar has to be regarded, in a very essential and pregnant sense, as a forerunner of Carnap's conception of a general logical syntax. One has 'only' to omit the detour through the realm of meaning, and the reliance upon an apodictic evidence and to add a mastery in modern symbolic logic and its philosophy in order to perform the transition from Husserl to Carnap" (Bar-Hillel 1957: 369). Gardiès (1975) very instructively develops the theoretical affinities between Chomsky and Husserl, and much closer to us, also the French phenomenologist J. Benoist concludes one of his numerous works on the Fourth *Logical Investigation* with the following remark: "One of the logical consequences of the considerations developed in the present work, particularly as regards the syntactic *a priori*, would certainly consist in readdressing the issue about the connection between Chomsky and phenomenology and about the possibility of developing a Chomskyan interpretation (minus mentalism) of phenomenology" (Benoist 1999: 268; my translation). And finally, D. Münch critically observes: "If we look at the fourth *Logical Investigation* we can see that Husserl is a forerunner of Chomsky. In this investigation Husserl applies the basic concepts of formal ontology which he had developed in the third investigation to grammar. Language is conceived as a field which is guided by purely formal rules, which has to be studied by a discipline he calls 'pure grammar' […] This grammar has an algorithmic character, too. But Husserl does not only support a computational approach to language. Moreover, it is one of his central claims that our intentions are restricted by the laws of purely logical grammar. Thoughts which are not in accordance with these laws are nonsense, i.e. they cannot be directed towards an intentional object" (Münch 2002: 203).

Now, the above is only partially true. The present claim is that Husserl in the Fourth *Logical Investigation* does not abstract *one*, but *two* grammatical a priori, which are not only essentially different, but also, in the framework of his analysis, incongruous with each other: on the one hand, we have a semantic-mereological a priori (valid for the coherent configuration of linguistic parts *qua* partial *significations* in a meaningful whole, and as such perfectly compatible with the characterization of intentional language use as developed in the First *LI*)), and, on the other, a genuinely syntactic a priori (valid for the linear combination of linguistic parts *qua tokens* of *syntactic categories* – notwithstanding their specific semantics). The first mereological-semantic a priori is developed in the first nine paragraphs of the Investigation, the second in the remaining. The first approach has *as such*[10] been ignored, neglected, dismissed as vague or considered as a mere preliminary exercise that only serves the function of introducing to the fundamental hypotheses of the last section.

In the following, we will invert the order of Husserl's Investigation and, thus, start with the syntactic a priori and end with a rehabilitation of the semantic-mereological a priori. To conclude we will see how research within Cognitive Linguistics has provided evidence that support the latter approach.

The Syntactic A Priori

Albeit short and without much import on phenomenological scholarship in general, the Fourth *Logical Investigation* exerted an immediate influence on an empirical science, modern, structural linguistics. Husserl's ambition was indeed to establish the foundations and conditions of possibility of an autonomous study of language and its objective correlate: language as such. It rested on the following fundamental claim:

> Modern grammar thinks it should build exclusively on psychology and other empirical sciences; As against this, we see that the old idea of a universal, or even of an a priori *grammar*, has unquestionably acquired a foundation and a definite sphere of validity, from our pointing out that there are a priori laws which determine the possible forms of meaning. (Fourth *Logical Investigation,* Introduction, p. 49)

This claim to the effect that language is a self-contained system, governed by its own laws, announces an exclusively linguistic or grammatical approach to the

---

[10]To avoid all possible misunderstandings I stress *as such*: the concepts used by Husserl in the first nine paragraphs (dependent and independent meanings, relations of foundation, etc.) neither have been ignored nor are they to be considered as essentially incompatible with syntactic analysis as such. As we have already seen, structural linguistics exploits exactly these conceptual tools in its characterization of language as an autonomous, functional system. And as Mulligan (1988) has shown there is nothing intrinsically inconsistent in a mereological approach to syntax: in his *Theory of Language*, Karl Bühler (1934) indeed exposes the basic tenets of such a syntactic theory (to a large extent developed in the very same terms by the founding father of modern theory of syntax, Lucien Tesnière (1959). My point is simply that no one seems to have further elaborated, even appreciated, Husserl's sketch of a comprehensive theory of *semantic* unity in mereological terms.

examination of (the possibility of) meaning construction in language and thereby defines the theoretical framework of a genuinely new science, structural or formal linguistics. Hence, what is examined in the Investigation, from §10 and on, is not how, by means of which *acts* people make use of language to express their meaning intentions, but how, by virtue of which *formal properties* language combines lexical entities into meaningful wholes. The combination of linguistic meaning is here considered notwithstanding its relation to or foundation on the intentional acts and their contents (for these belong to another material sphere with other characteristic material a priori laws).

The nature of the laws ruling the compounding of lexical entities is laid bare thanks to the fundamental and most well-know distinction of this Investigation: the distinction between *senselessness* (*Unsinn*) and *nonsense* (*Widersinn*); that is to say, between sequences of words that have no intrinsic meaning (for instance, 'formerly are walking or'), and sequences of words that do make sense whether or not they are true, formally consistent, or refer to any possible object of whatever sort (like 'stubborn quantifiers' or 'electric complex numbers'). The distinction serves the crucial purpose of clearly delimiting the genuine object of a "pure logical grammar." Indeed, however fundamental, logical laws establish truth (or falsity) only on the basis of intrinsically meaningful propositions. Before it can be considered true or false, a proposition has to make sense, and correlatively it does make sense even if it is flagrantly false. Therefore, the rules governing the compounding of partial significations and the rules governing the latter's formal consistency are not the same; rather, Husserl claims, there exist a priori laws, in the realm of pure significations, that establish the distinction between what makes sense and what does not make sense, and these "laws of sense […] *direct logic to the abstractly possible forms of meaning, whose objective value it then becomes its first task to determine*" (Fourth *Logical Investigation,* § 14, 71).

In other words, the distinctive feature of language is that only certain sequences of words make sense, while others don't. The task of the Fourth *Logical Investigation* (and linguistics in general) is in other words to define the laws governing the compounding of significations, "i.e. the a priori patterns in which meanings belonging to different semantic categories can be united to form one meaning, instead of producing chaotic nonsense" (Fourth *Logical Investigation,* § 1, 49). These laws would then make out a pure "morphology" (*Formenlehre*) of significations, in contradistinction to the pure logical theory of validity. Here follow the tenets of this morphology of meaning combination:

- What characterizes the combination of linguistic meanings is that they must be *connected* in specific, "antecedently determined" ways: only certain sequences of words make sense, others, even including the very same words, yield only a "heap of meanings, never a single meaning" (Fourth *Logical Investigation,* §10, 62).
- Since a word sign makes sense only insofar as it is part of a well-formed sequence of word signs, it is *dependent* on the compositional law supporting this sequence. Now, just as, in another sphere, the relation of mutual dependence between color and extension, is not one that concerns one specific quality and

one specific surface *qua*, say, exactly this red nuance and exactly this apple surface, but one that concerns the genera differences ('color,' 'extension') of which they are the lowest, the laws governing the combination of linguistic elements do not apply to the linguistic units *qua* specific meanings, but to the "essential genera" (*wesentlichen Gattungen*), i.e., the "semantic categories" (*Bedeutungskategorien*) to which they belong.[11] So, if a sentence like 'without cherished I alphabet an vowels' does not make sense, it is not due to the incompatibility of the particular semantics of these specific words, nor to the words themselves, but to the fact that these words as instances of specific 'semantic' categories do not combine according to the rules that govern the combination of semantic categories.

- Husserl's sketch of a pure logical grammar rides rather heavily on this notion of 'semantic' category. It is an upshot of the following fundamental principle: whenever we have a well-formed expression, say, 'This tree is green,' we can, by means of formal abstraction, obtain the corresponding pure "syntactic form"[12] '*S is p*.' Now, this ideal form can be instantiated in indefinitely many, but still evidently restricted ways. The variability is indefinitely rich insofar as any nominal matter – say, 'this carved square root,' 'this hungry bikini,' etc. – can instantiate *S* (or substitute 'this tree' in the former example) without affecting the grammatical integrity of the sentence; consequently, any adjectival matter – say 'agnostic' or 'skilled' – can instantiate *p* (or replace 'green' in the example). On the other hand, the variability is strongly constrained: a token of one categorical class can only be replaced by a token of the same categorical class. In 'this blue raven is green,' the integrity of the complex linguistic form remains invariant, whereas in 'this careless is green,' it has disintegrated due to an illegal exchange of syntactic categories and thus an illegal combination of categories which is not supported by any corresponding "pure form of meaning."

- The conclusion is now quite simple: a construction of any sort makes sense linguistically (notwithstanding its logical consistence) if and only if it is sustained by (or is an instantiation of) a pure syntactic form combining 'semantic' categories

---

[11] Cf., "The impossibility attaches, to be more precise, not to what is singular in the meanings to be combined, but to the essential *kinds*, the *semantic categories*, that they fall under" (Fourth *Logical Investigation,* 62/318). "Semantic category" translates "*Bedeutungskategorie*," but what is meant here is nothing but *syntactic* category. The fact that the same term, "*Bedeutung*" is used indistinctly throughout the whole Investigation (yet in two different senses) may have served to blur the essential differences between the two approaches.

[12] Findlay's translation does not faithfully render Husserl's expressions. In the German version, Husserl says "Gehen wir formalisierend von der gegebenen Bedeutung […] zur entsprechenden Bedeutungsgestalt, zur 'Satzform' über, so erhalten wir *dies S ist p*, eine Formidee, die in ihrem Umfang lauter selbständige Bedeutungen befaßt" (Fourth *Logical Investigation* (Husserliana Edition: 318–319). Findlay conflates "Bedeutungsgestalt" and "Satzform" into "pure form of meaning" (§10, 62), which contributes to obliterating the fact that whenever expressions such as "meaning" or "semantic (category)" are used in this and the following paragraphs, what is meant is syntactic 'function' or syntactic 'category.'

in a predetermined way. A word sign signifies something if and only if it is a variable of a category integrating a law-governed syntactic construction. The result of this is a research program for the *pure logical grammar* and its objective correlate – the "ideal scaffolding of language" – whose essential and foundational principle Husserl by now considers as evidently established:

> The task of an accomplished science of meanings would be to investigate the law-governed, essence-bound structure of meanings and the laws of combination and modification of meaning which depend upon these, also to reduce such laws to the least number of independent elementary laws. We should obviously also need to track down the primitive meaning-patterns and their inner structures, and, in connection with these, to fix the pure categories of meaning which circumscribe the sense and range of the indeterminates – the 'variables' in a sense exactly analogous to that of mathematics – that occur in such laws. (Fourth *Logical Investigation,* §13, 68)

The "ideal scaffolding" of language is thus something that a syntactic morphology is supposed to determine. It consists of a set of primitive syntactic templates which govern any actual compounding of lexical entities or from which any such compounding can be derived. The approach is obviously different from the one followed in the First *Logical Investigation*. Here, words came to mean something if and only if meaning was bestowed on them. No such perspective prevails in this section of the Fourth *Logical Investigation,* if language *can* make sense, and thus if users can confer any concrete meaning to it in the first place, it is because language as such is sustained by a formal *Bauplan* which prescribes certain forms of combination and proscribes others. Inversely in the First *Logical Investigation,* if language *does* make sense, it is because speakers use it with the specific purpose of giving shape to the contents of their mental acts. Language as dealt with in the First *Logical Investigation* is assessed by virtue of its being *essentially* a vehicle, serving the purpose of expressing and re-articulating already formed "ideas", "*Vorstellungen*" or "expressible thoughts," whereas in the second part of the Fourth *Logical Investigation* it is assessed as an autonomous, self-contained symbolic system, whose formal consistency is based on laws that apply on categorical forms, not on expressed meanings.

Both these a priori are valid. Speakers cannot randomly bestow meaning on just any acoustic matter; such matter must be phonologically and syntactically framed in order to be filled in semantically. Inversely, the phonological and syntactic scaffolding of language only provides us with possible meaning, sentences, but not with expressed thoughts (with conditions of satisfaction). All in all, so far we have two principles of organization, but no mutual articulation. One principle of organization operates on the configuration of intentional contents; the other, essentially different one, operates on categorical forms within a symbolic system. A crucial task therefore remains: if the ideal scaffolding, thus defined, is purely syntactic it cannot explain how language expresses thought, ideas, contents, and thus how the contents and structure of intentional acts can be transposed and faithfully represented in language. We therefore need an intermediary level of configuration which can account for those structures and forms by virtue of which language can shape and represent the contents of our intentional experiences.

This is exactly what Husserl set out to establish in the first part of the Fourth *Logical Investigation*. He does so in terms of a semantic-mereological a priori which in many respects is radically different from the syntactic a priori above. It lays bare a level of representation, the level of conceptual structure, where language and perception/intentional acts are compatible, i.e. where intentional contents can be matched with *semantic* forms, and where language, through these semantic forms, can consistently evoke intentional contents.

The Semantic-Mereological A Priori: What Governs the Configuration of Partial Meanings into Signifying Wholes?

One of the fundamental things we ask of a "pure logical grammar" is to determine the principles that govern the combination of linguistic units into a whole. Now, this issue can be addressed syntactically, as we just have seen, and the issue is then addressed in terms of the a priori patterns governing the *linear* combination of linguistic parts into well-formed strings of syntactic categories (or ideal syntactic templates). However – and even though linguistic theory for decades has neglected this – the issue can also be addressed semantically, and the guiding question is then the following: given that a complex linguistic object (say, a sentence or a compounded expression) is made out of signifying parts, are there any rules that govern the combination of these parts, not by virtue of their syntactical or categorical form, but by virtue of their signification? It is indeed this line Husserl chooses to follow in the first 9 paragraphs of the Fourth *Logical Investigation,* whereby he outlines a theory of linguistic composition and compositionality on semantic grounds.

As it is developed, the question is mereological: it concerns the relations between signifying parts and their whole; it is semantic, of course, because the task consists in characterizing the way in which each part participates in the expression or rearticulation of a full representation. Therefore, Husserl proceeds by examining the order of complex linguistic objects in terms of the general part-whole theory developed in the Third *Logical Investigation*, and more specifically his first step consists in determining whether linguistic parts are all of the same type, or whether the distinction between *independent* and *dependent* parts also obtains in the case of linguistic entities, the one Husserl – following C. Stumpf – found to hold generally between parts of a whole in the Third *Logical Investigation.*

As is well known, independent parts (or *Stücke*, pieces) are such that they can be considered in isolation, whereas dependent parts (or *moments*) must necessarily occur with other parts of a certain sort in a whole of a determinate sort. Thus, a horse's head, its tail, its left front leg are such *detachable* pieces, whereas its color and shape are *non-detachable* moments of the whole (no color without extension and vice versa). In this respect *moments* of a whole play a fundamental role since they are necessarily related to other parts of a determinate sort in a whole of a determinate sort – they are *founded* on them in relations of either unilateral or mutual dependence: they are, as it

were, crystallizations of the system of relations that bind the parts into one consistent whole.[13] This is what Husserl applies to language and meaning: granted that linguistic objects, word-complexes, are made out of different significant parts, we can now ask whether all parts have the same *kind* of signification (in this case sentences are summative wholes of significant parts) or whether it is possible to distinguish between different types of signification, just as it is possible to distinguish between dependent and independent contents, moments and pieces (in which case, sentences or complex linguistic objects are internally articulated in virtue of determinate types of relations that it would be possible to characterize)?

Husserl endorses the latter view and proceeds to a three-step determination of what is meant by "dependence" in the realm of linguistic significations. First he draws an exclusively *grammatical* distinction between two general types of linguistic parts in discourse; then he leads this difference back to a genuinely *semantic* distinction; and ultimately he founds the linguistic semantic distinction on a *cognitive* distinction, i.e. a general distinction between types of meaningful contents given in intentional acts. Here is how the argument goes:

To begin with (Fourth *Logical Investigation*, §4), Husserl refers to the medieval parsing of the elements of discourse into two general subsystems: *categorematic and syncategorematic* expressions. The distinction is originally purely grammatical: sentences are traditionally said to be compounded by words that have a proper signification (generally, nouns, verbs, and adjectives were considered as such), called *categorematica*, and words (like prepositions, conjunctions, particles, pronouns, as well as grammatical units such as inflexions, prefixes, suffixes, etc.) that only take on a signification together with other words; the latter are called *syncategorematica*, or co-signifying words. Husserl now presents the following alternative: (1) The distinction is indeed merely grammatical, i.e., only relative to the particular make-up of each empirical language. In that case, only categorematic words are linked to representations proper (they express the latter), while syncategorematic words merely play a role at a grammatical level, i.e., they only have a function as connectors within the linguistic system of expression, but they do not express any meaning whatsoever. If so, "the syncategorematic words which help to build up [the] expression are, properly speaking, quite meaningless, only the whole expression really has a meaning" (Fourth *Logical Investigation,* §4, 54). (2) The completeness or incompleteness, i.e. the independence or dependence, of a linguistic expression (a word sign) is a reflection of a completeness or incompleteness at an underlying semantic level. In this case, *syncategorematica* do not simply play a functional-connecting role at a grammatical level of expression, they have a signification in their own right, yet their signification is not a complete or *independent* one, but an incomplete and *dependent* one.

Husserl emphatically adopts the latter solution: *categorematica* are *semantic pieces* of a linguistic whole, *syncategorematica* are *semantic moments* of a linguistic whole.

---

[13] This is why moments, though dependent, are said to be *prior* to pieces: "Strictly speaking our approach is positive in the case of what is dependent, negative in the case of what is independent" (Third *Logical Investigation,* §7, 13): moments are by essence structural parts.

The grammatical distinction is an expression of an essential semantic distinction, strictly analogous to the distinction between dependent and independent contents of intentional acts. The answer is adopted on the grounds of the following claims, which show not only the comprehensive character of Husserl's approach (it addresses the essence of its object, language, at three distinct and interrelated levels: a grammatical, a semantic, and a cognitive one), but also the kind of mereological principle governing the unity of complex linguistic objects that he aims at laying bare. Let us consider the two fundamental assumptions.

The first claim follows of course directly from the attribution of dependent significations to syncategorematic expressions: linguistic complex objects are simply tokens of complex objects in general, thus the laws governing their unity are the same as those established within the general theory of wholes and parts. Parts are combined with each other by virtue of their essential dependent or independent contents, i.e., by virtue of the type of connection to other parts they require. The a priori governing the realm of linguistic significations is therefore a *mereological* a priori:

> Having called syncategorematic meanings 'dependent', we have already said where we think the essence of such meanings lies. In our enquiries into dependent contents in general, we have given a general determination of the concept of dependence: it is this same dependence that we have to recognize in the field of meaning; Dependent contents, we stated above (Third *Logical Investigation,* §§5–7), are contents not able to exist alone, but only as parts of more comprehensive wholes. This inability [of dependent significations to exist alone] has its a priori governing ground in the specific essences of the contents in question. Each dependence points to a law to the effect that a content of the sort in question, e.g. sort α can exist only in the context of a whole $G(\alpha\beta ... \mu)$, where $\beta ... \mu$ stand for *determinate* sorts of content. 'Determinate', we said, since no law merely asserts connection between the sort α and any [*arbitrary*] sort of [context] […] Law involves specific determinateness of context: dependent and independent variables have spheres limited by fixed generic or specific characters. We have mainly employed as examples the concrete things of sensuous intuition. We could, however, have brought in other fields, those of act-experiences and their abstract contents. (Fourth *Logical Investigation,* §7, 58–59)

The argument is conducted from an entirely a priori point of view and, at this level, it is quite clear. To the extent that we can identify dependent parts within a linguistic whole, and to the extent that such parts call for completion, and not just any completion, but of a *determinate* kind, there must exist laws governing the *way* in which such parts should be completed, with what kind of parts, and in what types of connection. That is to say, just like dependent contents in general, dependent significations are essentially prior or more fundamental than independent significations, since they point at or imply the laws governing their completion (cf. Fourth *Logical Investigation,* §7, 58).

However, if we remember that here Husserl is engaged in a linguistic inquiry, we may be entitled to ask for specifications of this semantic scaffolding of language: what does it mean for a *linguistic content*, by virtue of its *specific essence*, to require a context of a determinate sort? In fact, it should be easy to illustrate this point within the linguistic domain in view of the fact that in the Third *Logical Investigation* Husserl provided enlightening examples of what he meant by dependent content, relations of foundation, and relations of unilateral or mutual dependence

within the domain of intuition. Yet, in the Fourth *Logical Investigation,* there is *no* linguistic counterpart to, say, the relation between color and extension such as it is minutely analyzed in the preceding Investigation. In other words, it is quite difficult to tell how exactly the laws invoked apply to the domain of linguistic significations, and, certainly, how laws governing the relation between contents of intuition should rule the unity of linguistic parts.

If we now turn to the second claim underpinning Husserl's argument, we will see that it in fact clearly specifies what is meant by linguistic "content" and its "specific essence." The main tenet of the argument is *functional*. Language serves the purpose of expressing representations or experiences (whether actually on-going, recalled or imagined). Now, experiences are themselves made out of a great many partial experiences, combined into a whole or intended as connected in a specific way. Husserl's point here is simply that if language had not disposed of the necessary means to express both partial experiences or representations *and* the specific way in which these are intended to combine, it would be incapable of expression, and thus not be language:

> Language has not been lead by chance or caprice to express presentations[14] with names involving many words, but by the need to express suitably a plurality of mutually cohering part-presentations, and dependent presentational forms, within the enclosed self-sufficiency of a presentational unity. Even a dependent moment, an intentional form of combination through which, e.g., two presentations unite in a third, can find semantic expression, it can determine the peculiar meaning-intention of a word or complex of words. Clearly, we may say that if presentations, expressible thoughts of any sort whatever, are to have their faithful reflections in the sphere of meaning-intentions, then there must be a semantic form which corresponds to each presentational form. This is in fact an a priori truth. And if the verbal resources of language are to be a faithful mirror of all meanings possible a priori, then language must have grammatical forms at its disposal which give distinct expression, i.e. sensibly distinct symbolization, to all distinguishable meaning-forms. (Fourth *Logical Investigation,* §4, 54–55)

It is worthwhile to stress once again that the approach is semantic through and through. The unity of complex linguistic wholes is defined, not in terms of syntactic constituency, but in terms of genuinely semantic relations of foundations between dependent significations and those specific contexts of meaning in which they must occur.[15] What is more, in the vein of the First *Logical Investigation*, the approach is

---

[14] "Presentation" translates *Vorstellung* (idea, representation).

[15] In view of the entirely syntactic analyses, which follow in the Investigation, this statement is indeed astonishing: it is easily seen that Husserl's mereological foundation of linguistic constituency is not at all applicable to syntax in his sense. Let us consider the syncategorematic expression α as a syntactic category – regardless of any signification – ; we now apply the law according to which it should appear in a (syntactic) whole of a specific sort $G(\alpha\beta \ldots \mu)$, where $\beta \ldots \mu$ stand for *determinate* sorts of syntactic categories. Obviously the extension of this law (say, the number of syntactic constructions in which α, e.g., the conjunction 'but,' can occur) is indefinitely big, and thus quite indeterminate. Now, compare this to its *semantic* counterpart, where 'but' *expresses* an "intentional form of combination;" in this case, "but" requires only one *sort* of context: one in which a relation of contrast, tension, or conflict is intended to hold between two partial representations combined by 'but.'

semantic *twice* rather than once: first, syncategorematic expressions have genuine significations (which must ultimately imply that these are amenable to semantic description); next, the significations are themselves *correlates* to certain pre-linguistic contents of meaning acts that in turn constitute a complex, configured whole of partial meaning acts, so that the dependence or independence of each signification is a result of its capacity or incapacity to be the "full, entire meaning of a concrete act of meaning." In short, as Husserl concludes, a linguistic (dependent or independent) content and its specific content is an articulation of the content of a *pre-linguistic* meaning act. Syncategorematic expressions are thus linguistic forms that *specify grammatically* the content inherent in a specific mode of intention (or, as quoted above: they are grammatical forms that express "intentional forms of combination" of partial representations):

> A meaning, accordingly, may be called 'independent' when it can constitute the *full, entire meaning of a concrete act of meaning*, dependent, when this is not the case. It can then only be realized in a dependent part-act in a concrete act of meaning, it can only achieve concreteness in relation to certain other complementary meanings, it can only exist in a meaningful whole. The dependence of meaning *qua* meaning thus defined determines, in our view, the essence of the syncategorematica. (Fourth *Logical Investigation,* § 7, 59)

We may now sum up Husserl's claims concerning the unity of meaning *qua* meaning and contrast it with it the determination of linguistic unity in terms of syntactic constituency:

- Husserl sets out from the purely *grammatical* distinction between categorematic and syncategorematic expressions.
- He founds the grammatical distinction on a *semantic* distinction between independent and dependent significations. So doing, he asserts the existence of two general subsystems in language that each contribute its *type* of meaning. The categorematic (or lexical) subsystem contributes independent significations that can be apprehended per se; the syncategorematic (or grammatical) one contributes significations that are unbounded, 'vague,' and "call for completion." Due to the laws that govern the configurations of parts into wholes, independent parts (or 'moments') require not just any whole whatsoever, for they are not amenable to all sorts of completion; rather they require completion of a specific sort. Although Husserl does not conclude this himself, it follows from the above that the meaning contributed by syncategorematic expressions *is the general semantic frame or semantic structure within which they are to appear*.
- Moreover, linguistic significations are defined as correlates of meanings intended in specific meaning acts. Dependent and independent significations are counterparts of dependent and independent contents of meaning acts, so that the structure governing the combination of significations at a *linguistic* level is a structure already ruling the configuration of *pre-linguistic* intentional acts. Thus, syncategorematic expressions are not merely grammatical or syntactic connectors, they are "faithful" "reflections" of "intentional forms of combination" through which partial representations unite into one complex representation.
- The object 'language as such,' which is assessed in this functional approach, is *not* accessed as a self-contained, autonomous object, i.e. by virtue of its specific

essence *qua* that kind of object, but rather by virtue of its being an object whose *essential* function is to be a *symbolic vehicle*, a means to express, faithfully reflect and rearticulate already formed, structured, or configured pre-linguistic contents of meaning acts.

- While both approaches, the syntactic and the semantic, invoke "laws of combination", in the former these are specified as laws of *connectivity* governing the principled *sequentialization of parts*, whereas the latter specifies them as laws governing the *configuration of parts* by virtue of their meaning.

- Moreover, both approaches refer indistinctly to mereological principles of combination in terms of the kind of relations implied by either dependent and independent parts; this is probably the reason why quite few scholars have noticed the very existence of these two approaches (as shown in detail in Bundgaard 2004a, 61ff. the transition or *metabasis* occurs in §10, 62ff.). In other words the conceptual framework is the same (or the conceptual wording is identical). However, as the categorical content of the concept "laws" or "form" of combination changes from one defined as "intentional", and thus pre-linguistic, to one construed as a purely linguistic connector, independently of any intentional acts sustaining any use of words, then, as a consequence, the dependence/independence distinction applies differently in either cases. In fact, in the syntactic approach this distinction is, as concerns the parts themselves, obsolete: since a part is significant only insofar it appears in a well-formed sentence derived from an ideal syntactic template, then *all* parts are equally dependent on their whole. Very differently so in the semantic approach where the unity of the whole is claimed to rely on specific dependent meanings, which therefore could be defined as *figural* or *unifying* moments, i.e. parts that do not only specify an element of a semantic whole, but also evoke this whole and its structure as such.

- Finally, Husserl's wording is misleading in the sense that it, too, blurs the difference between the two approaches: when in the first part of the *Investigation* he talks about signification ("*Bedeutungsintention*" (meaning-intention), "*Form auf der Bedeutungsseite*" (semantic form), as in §4), then it is signification proper he is referring to. Yet, when, from §10 and on, he starts assigning a crucial role to the ideal elements referred to as "semantic categories" ("*Bedeutungskategorien*"), these are to be construed as *syntactic* categories.

So much for the tenets of and differences between each of these approaches, but where does this lead us? From a strictly exegetical point of view: hardly anywhere. Husserl himself seems to have abandoned the idea of a semantic-conceptual definition of the unity of linguistic meaning as soon as he had developed it, not only because he redefines the whole framework of the examination (around §10 in the Fourth *Logical Investigation*) and leads his description of language as such in a decidedly syntactic direction, but also more fundamentally: he never picked it up again (with the partial exception of the analyses from *Experience and Judgment*, Husserl 1939). On top of this, Husserl provides absolutely no systematic evidence for the fact that "syncategorematic expressions are understood, even when they occur in isolation; they are felt to carry definite 'moments' of meaning-content, 'moments' that look forward to a certain completion which, though it may be

indeterminate materially, is formally determined together with the content in question, and is circumscribed and governed by it" (Fourth *Logical Investigation*, 56/301). He does indeed claim that "any example proves" this as well as the point that syncategorematica express intentional forms of combination, but no linguistic example in particular is recruited to corroborate this assertion. In other words, no concrete analyses illustrate what the "semantic forms" are which correspond to the contents of the intentional acts, nor how one should concretely construe the decisive configurational function of dependent significant parts. The absence of examples has obviously had serious drawbacks for the argument, since no one seems to have shared Husserl's confidence in the self-evidence of his point. None of the linguistic theories that have either explicitly acknowledged their debts to Husserl or been indirectly influenced by him have elaborated these semantic aspects of his considerations. On the contrary, at least until the beginning of the 1980s, they have remained a sort of *terra obscura* for linguistic theory. This is, as we shall now see, far from being the case any longer.

Language and Conceptual Structure – Evidence from Cognitive Linguistics

In its general progression and development, linguistics seems to confirm the importance of Husserl's analysis of language. Most importantly, since the late 1970s a whole research program, Cognitive Linguistics, has developed whose main goal is indeed to establish and systematically map the nature of the semantic forms which correspond to, evoke, and linguistically specify the form of our intentional acts and their contents. Cognitive linguists have – independently of Husserl – (a) unfolded a theory of the semantic-mereological design properties of language which show striking analogies to Husserl's Fourth *Logical Investigation*; (b) provided systematic and rich evidence for what exactly should be understood by "dependent" significations which "call for completion" or require a whole of a determinate sort; and (c) established how language can specify and express our mental act contents (the partial representations *and* the "intentional forms of combination") in terms of such dependent significations. This is, of course, not the place to introduce thoroughly to Cognitive Linguistics as such, suffice it, as a conclusion, to give three examples of how it has supplied with insight in the semantic relation between mind and language according to the three aspects just listed.

Len Talmy's Closed Class Semantics

As regards the design properties of language, Len Talmy has in his work (Talmy 2000) also called attention to the well-known feature that language is composed by two subsystems, a *grammatical* and a *lexical* one. The lexical subsystem (matching

Husserl's *categoremata*) is composed by open word classes (nouns, verbs, adjectives: classes which are rich in member and likely to undergo changes), whereas the grammatical subsystem (matching Husserl's *syncategoremata*) integrates closed word classes (all other word classes – prepositions, conjunctions, verb particles, but also tense and case markers, etc., i.e. classes which have only few members and are not readily subject to any augmentation or change). This is commonplace in linguistics, but what is less traditional, in fact quite unique, is that Talmy, exactly as Husserl, attaches different semantic functions to each of these classes. In the very same vein as Husserl, Talmy first considers that complex linguistic expressions (say, a sentence) serve the function of "evoking in the listener" a specific "experiential complex" or a "cognitive representation" (Talmy 2000, p. 22). Next, he remarks that elements from both subsystems represent "conceptual material" – they all have a signification – but not the same type of material. Keeping in mind that language according to Husserl must dispose of semantic forms to express not only bounded contents of meaning acts, but also the way these are intended to relate (their "intentional form of combination"), it is easy to appreciate the affinities to Talmy's conception:

> The grammatical and lexical subsystems in a sentence seem generally to specify different portions of a cognitive representation (CR). Together, the grammatical elements of a sentence determine the majority of the *structure* of the CR, while the lexical elements together contribute the majority of its *contents*. The grammatical specifications in a sentence, thus, provide a conceptual framework or, imagistically, a skeletal structure or scaffolding for the conceptual material that is lexically specified. (Talmy 2000, p. 21)

Such assumptions, as well as Husserl's in his semantic-mereological approach, clearly call into question the task assignment dogma within linguistics, according to which perception provides referential access to objects, whereas language combines the symbols referring to these representations exclusively in terms of its own laws of linear composition. (A position to which Husserl himself provided the theoretical foundation in his syntactic approach.) It claims, on the contrary, that linguistic representation is not a two-relata business between linguistic structure proper and reference objects, but a three-relata business, so that language, thanks to certain expressive and semantic properties, not only refers to a scene, but also articulates structured modes of *experiencing* or *conceptualizing* the referent scene. The crucial claim here is then strictly like Husserl's: closed-class elements play a privileged role (as dependent meanings) since they make out the inventory of semantic forms through which the structures of a complex representation are specified, or, to put it differently, through which a referent scene is expressed linguistically such as it is intended in experience. Talmy thus concludes that a study of semantics as a genuine combinatorial system in language must be a systematic study of the kind of structure specified by closed-class elements, since they are in charge of the "conceptual organization" within language. Talmy's monumental work within cognitive semantics can be considered a detailed investigation and mapping of "the conceptual structuring system of language" (Talmy 2000, p. 21).

Fillmore's Frame Semantics

Charles Fillmore (1982) describes in his frame semantics a vast array of cases where the understanding and thus the signification of a lexical entity implies the activation of a network of correlated terms together with which the entity forms a structured whole or with respect to which the named entity takes on a specific meaning. The correlated or necessarily co-intended network of terms is what Fillmore calls the frame of a term. Thus, terms as "diameter," "hypotenuse," "quarterback," and "Wednesday" are clearly framed, and thus clearly dependent significations, to the extent that they only take on a meaning on the backdrop of the whole relative to which they are defined. The same goes for actions that integrate ordered event types (or "scripts" as Shank and Abeles had it) such as "pay," "sell," etc. which evoke the well-known "commercial" frame. As is manifested from the above standard examples semantic dependence clearly extends beyond the class of syncategorematic terms; however, the latter are still epitomies of lexical entities whose meaning is in need of completion and thus structural. Consider

1a. They are married, and they don't live together
1b. They are married, but they don't live together

Both sentences have same reference, they intend the same object (two persons are married, two persons don't live together), but they do so through different meanings, thus constituting different states of affairs. The difference is, evidently, relative to the conjunctions, syncategorematica with dependent significations. Several things should be noticed. Firstly, the conjunction does not simply express an additional property of the objective scene referred to; the meaning of the conjunction affects the whole global expression (the meanings of 1a and 1b are globally, not locally different). Secondly this shows, in a flatfooted way, what it means for a dependent signification to require a whole of a determinate sort. The meaning of "but" could be characterized as follows: in a complex construction compounded with "but," whatever is to the left of "but" and whatever is to the right of it are intended as "contrasting" or "conflicting" contents in some respect; they take on this crucial, semantic value by virtue of the dependent content that combines them. Therefore dependent contents do in fact not simply require determinate contexts; they *contribute* the kind of semantic whole into which the partial significations are combined (in the same sense that "hypotenuse" and "sell" contribute a specific frame). This semantic whole could in the case of "but" be diagrammed as a schematic function (in Frege's sense) like this: $X > [but] < Y$.[16] Thirdly, and crucially, the schematic meaning of the conjunction expresses an intentional form of combination, that is to

---

[16] Notice that this also holds for "and": when "and" applies to a multiplicity it may very well express that the multiplicity is intended or conceptualized as a group. Thus "disco, Wittgenstein, *Fingerspitzgefühl*" is a multiplicity, whereas "disco, Wittgenstein, and *Fingerspitzgefühl*" is (somehow intended as) a group; for example because the second letter in each word is *i*.

say it expresses how speaker conceptualizes and assesses the relation between the partial representations (they are married/they don't live together). In other words it illustrates precisely that property of language which Husserl in the above quotation (cf. section on The Semantic-Mereological A Priori: What Governs the Configuration of Partial Meanings into Signifying Wholes) considered as essential: the schematic meaning of, say, "but" is the semantic form which corresponds to, faithfully expresses and grammatically specifies the mental acts of the language user.

This brings us to the third case.

Grammatical Specification of Perceptual Intentionality

In general, cognitive linguists have convincingly shown how language is capable of grammatically specifying the *act matter* (in Husserl's sense) of an intentional experience and thus the way an object has been perceptually intended. This is manifest in cases of *alternations in schematization*; that is to say cases where different expressions refer to same object/scene: the variations on the linguistic level then reflect variations in the perceptual organization of the experience, the mode of givenness of the object. Such cases evidently show that language does not only name an object, it also specifies the character of the intentional relation to the object. Or in other words: the matter of the intentional act specified by the linguistic expression has both an objective and a subjective aspectual shape: the former concerns that property of the object which is profiled or that aspect under which it is considered; the latter concerns the intentional subject's own relation to the object (both epistemically as in the above "but" example and perceptually as in the examples below): both the objective and the subjective aspect of the act matter have a linguistic counterpart. The following examples are just a small selection of cases which show how basic intentional/cognitive distinctions are expressed by language as a full-fledged part of the articulated meaning:[17]

- (*Mode of perceptual processing,* specified by dependent significations) 1a. There are some houses in the valley ("synoptic" scanning) vs. 1b. There is a house every now and then through the valley ("sequential" scanning).[18]

---

[17] Talmy (2000), Langacker (1987) have quite systematically mapped the correlations between mode of perceptual experience and linguistic specification.

[18] Cf. Talmy (2000, p. 71). Langacker (1991, p. 78 sq.) operates with the same distinction between "summary" and "sequential scanning". Langacker notices that such differences in perceptual intentionality (what he calls the "construal relation" to the object) can be expressed by the very grammatical/categorical form of the expression. "Sequential scannings" are likely to be expressed by finite verbs, whereas "summary scanning" are likely to be expressed by means of prepositions or nominalizations of verbs. So even though "the apple falls to the ground" and "the falling apple" share the same reference scene, they express different types of conceptualization by means of different kinds of grammatical categories each of which have semantic import. The idea that

- (*Profiling; figure/ground carving*, specified syntactically) 2a. *x* resembles *y* vs. b. *y* resembles *x* vs. 2c. *x* and *y* resemble each other.
- (*Point of view*, specified by the schematic meaning of prepositions) 3a. The boat is on the water (distal point of view; water idealized as a surface) vs. 3b. The boat is in the water (proximal point of view, water idealized as a container).
- This evidence from Cognitive Linguistics does of course not pretend to be either systematic nor exhaustive. Its only ambition is: (a) to immediately illustrate what semantic function dependent significations may fulfill; (b) to show how mental acts can be bound up with linguistic expressions; (c) and thus to show how intentional forms of combination can be specified by semantic forms/dependent significations. Incidentally, it also aims to show that the whole aspect of Husserl's Fourth *Logical Investigation* which has been generally neglected by both Husserl scholars and linguists in the over seventy years that followed its publication, has been overtaken and, in fact, systematically further elaborated by cognitive linguists. This revival has taken place independently of Husserl's theory of language. However, the fact that it picks up core features of the Husserlian analysis is not coincidental insofar as cognitive linguistics explicitly aims at establishing exactly the kind of the junction which was also at the heart of the Husserlian approach to language: to wind up linguistic structures with mental acts, while laying down the conceptual structure which makes out the principled fundament of this junction.

Now, this attempt to embed language in cognition or linguistic signification in intentional meaning has a crucial consequence which may, at least partially, explain why Husserl deviated from it around § 10 in the Fourth *Logical Investigation*, in principle devoted to the determination of the ideal structure of language as such: namely, that semantics is not linguistic proper.[19] Meaning is of course evoked, expressed, specified in language, by virtue of a set of determinate properties of the language system, but it clearly exceeds the domain of language in the sense that language can be shown to give shape to already organized, pre-linguistic, intentional contents of experience. If the key issue is how language can faithfully express our representations and their intentional form of combination, then, obviously, meaning cannot be restricted to language proper and explained solely in linguistic

---

grammatical categories do have semantic import is key to cognitive linguistics and indeed also to phenomenological linguistics. Roman Ingarden (1931) made exactly the same point when he established that the difference between verbal and nominal significations/expressions stems from the *type of intentionality* implied by their categorical contents, i.e. from the kind of "construal relation" to the object they express *qua* verbal or nominal (cf. Ingarden 1931, p. 76 ff.).

[19] If the issue of the Fourth *Logical Investigation* is language as such and if language is a phenomenon endowed with certain properties and characteristics that make it evidently distinguishable from other objects, then the essential distinctions proper to language should, allegedly, be drawn within the sphere of grammar itself. The essential properties assigned to language should be extracted from language itself, and not from the relation between linguistic forms of meaning composition and intentional forms of combination. Hence the reorientation effectuated by Husserl in the middle of the Fourth *Logical Investigation.*

terms: the syntactic approach developed in the second part of the Fourth *Logical Investigation* tells us how language as such may combine syntactic categories, not how it configures significations.

Therefore neither Cognitive Linguistics nor Husserl's theory of language are *linguistic* theories proper (which does not imply that they have no import on linguistics). Rather, they are comprehensive cognitive theories that examine and lay bare (1) the relative dependence of linguistic structure on pre-linguistic structure; (2) the essential tenets of pre-linguistic structure; and (3) the design features of the linguistic system that make it capable of systematically expressing and re-articulating the conceptual structure thanks to which we shape our experiences of objects.

Conclusion

However vague, incomplete or even heterogeneous Husserl's contributions to a theory of language may have remained, he nevertheless approaches language under the three fundamental aspects according to which most subsequent theory and philosophy of language have determined their object: the constituents of language use, the formal laws determining the structure of the vehicle, and the material laws governing the order of the contents represented by the vehicle.

In Husserl, we thus find the outline of three approaches to language:

• There is an *intentional* approach, which accounts for the ontology and the phenomenology of language use; i.e. the types of acts essentially and circum-stantially involved in language use, and the types of mental acts involved in the actual meaning construction and communication effectuated by means of lan-guage use – this approach is pre-linguistic and developed in the First *Logical Investigation*. It is an obvious forerunner of both linguistic and philosophical pragmatics (the latter in the Austin-Searlean vein).

• There is a *syntactic* approach, which accounts for the principles ruling the well-formed combination of linguistic units, i.e. the principled generation of well-formed sentences – and thus possible meaning – on the grounds of ideal syntactic templates; this approach is purely intra-linguistic and developed in the second part of the Fourth *Logical Investigation* (from § 10 and on). It is an obvious forerunner of orthodox Chomskyan linguistics and formal grammars.

• In between these two approaches, there is the *semantic* approach, which claims that the mereological properties of the contents of our intentional acts can be faithfully reflected linguistically: it is a design feature of language that it is composed by two sub-classes, one of which is specialized in expressing partial representations (*categorematica* and their independent significations), while the other (*syncategorematica* and their dependent significations) is specialized in expressing relations between partial representations (a) such as these relations hold between the objects referred to in and through the representations, and (b) such as speaker has intended these relations. This approach is cross-modal with

respect to the intentional domain of mental acts and the domain of linguistic meaning construction and is developed in the first nine paragraphs of the Fourth *Logical Investigation.* It is an obvious forerunner of Cognitive Linguistics.

The semantic-mereological approach is of decisive importance both within a Husserlian framework and in general. For if we go on taking the Pragmatics/Formal Grammar Yalta for granted, we seem to dispose of two essentially different and internally unrelated determinations of meaning in language: (a) signification in language obtains when meaning is intentionally bestowed upon a word sign; (b) signification in language obtains when a word sign appears as a variable of a category in an ideal syntactic template (a construction). These determinations must somehow be articulated since if language use is bound up with mental acts, then, inversely, it must be possible to determine the genuine linguistic forms onto which mental acts can, in effect, be hooked. The semantic-mereological level of language can be shown to play this mediating function, i.e. be the bridge locus between language taken as a structured symbolic form system and experience taken as a structured intentional form system.

References

Bar-Hillel Y (1957) Husserl's conception of a purely logical grammar. Philos Phenomenol Res 17/3:362–369

Benoist J (1999) L'A priori conceptuel – Bolzano, Husserl, Schlick. Vrin, Paris

Benoist J (2001) Intentionalité et langage dans les Recherches Logiques de Husserl. PUF, Paris

Benoist J (2002) Entre acte et sens – La théorie phénoménologique de la signification. Vrin, Paris

Bühler K (1934) Theory of language – the representational function of language. John Benjamins, Amsterdam/Philadelphia (1990)

Bundgaard PF (2004a) The ideal scaffolding of language: Husserl's fourth logical investigation in the light of cognitive linguistics. Phenomenol Cogn Sci 3:49–80

Bundgaard PF (2004b) Configuration sémantique et combinaison syntaxique dans la IVème Recherhe logique de Husserl. Recherches husserliennes 21:3–33

Saussure F.de 1915 Cours de linguistique générale. Payot, Paris, 1960

Fillmore Ch. J (1982) Frame semantics. In: The Linguistic Society of Korea (ed) Linguistic in the morning calm. Hanshin Publishing Co, Seoul

Gardiès J-L (1975) Esquisse d'une grammaire rationnelle. Vrin, Paris

Husserl E (1890) Zur Logik der Zeichen. In Husserliana 12. Nijhoff, Den Haag 1970, pp 340–373

Husserl E (1900–1901) Logische Untersuchungen, 2nd edition. Max Niemayer Verlag, Tübingen. (1913/1980). English translation: Logical investigations. Routledge, London/New York, 2001

Husserl E (1929) Formale und transzendentale Logik. Max Niemayer Verlag, Tübingen, 1981

Husserl E (1939) Erfahrung und Urteil. Felix Meiner Verlag, Hamburg, 1985

Ingarden R (1931) Das literarische Kunstwerk. Max Niemayer Verlag, Tübingen, 1972

Langacker R (1987–1991) Foundations of cognitive grammar, vols. I–II. Stanford University Press, Stanford, CA

Langacker R (1991) Concept, image, and symbol. The cognitive basis of grammar. Mouton de Gruyter, Berlin/New York

Mulligan K (1988) On structure – Bühler's linguistic and psychological examples. In: Eschenback A (ed) Karl Bühler's theory of language. John Benjamins, Amsterdam, PA, pp 203–226

Münch D (2002) The relation of Husserl's logical investigations to descriptive psychology and cognitive science. In: Stjernfelt F, Zahavi D (eds) 100 years of phenomenology – logical investigations revisited. Kluwer, Dordrecht

Smith B (1987) Husserl, language, and the ontology of the act. In: Buzzetti D, Ferriani M (eds) Speculative grammar, universal grammar, and philosophical analysis of language. John Benjamins, Amsterdam, pp 205–227

Talmy L (2000) Toward a cognitive semantics. MIT Press, Cambridge, MA

Tesnière L (1959) Eléments de syntaxe structurale. Kliencksieck, Paris, 1982

Metaphor and Cognition

Mark Johnson

Over the past half century metaphor has moved from being a peripheral topic in the philosophy of art to the status of a major philosophical, linguistic, psychological, and scientific issue in the theory of cognition. For over 2,000 years prior to this dramatic change of status, metaphor was regarded as nothing more than a figure of speech serving various rhetorical purposes, but not an essential part of human thought. In sharp contrast, contemporary empirical research on language and cognition affords metaphor pride of place at the center of abstract conceptualization and reasoning. This new view appears to require a serious reconsideration of some of the grounding assumptions of mainstream philosophy of mind and language.

The Traditional View of Metaphor

Aristotle (1941) is typically credited with providing the first systematic treatment of the subject and thereby establishing a view that has persisted, in one version or another, down to the present day:

> Metaphor consists in giving the thing a name that belongs to something else; the transference being either from genus to species, from species to genus, or from species to species, or on grounds of analogy. (*Poetics* 1457b).

Certain key features of Aristotle's definition eventually came to define central tenets of the received view that subsequently influenced both commonsense views and theoretical perspectives in a profound way. First, metaphor is treated as a linguistic matter – a use of words or other symbols – and not a key process of cognition that shapes our basic concepts. Second, metaphor is thought of as appropriating a name that properly belongs to one thing and applying it deviantly to some other thing. In more recent times, this view has come to be interpreted as the idea that so-called "literal" terms are the proper medium for picking out and describing aspects of our

M. Johnson
Department of Philosophy, University of Oregon

experience, so that metaphor becomes a deviation from ordinary literal sense. Aristotle's third key idea is that there are cognitive principles underlying metaphoric transference of meaning, the most prominent of which is similarity: "the greatest thing by far is to be a master of metaphor. It is the one thing that cannot be learnt from others; and it is also a sign of genius, since a good metaphor implies an intuitive perception of the similarity in dissimilars" (*Poetics* 1459a), and "The simile is also a metaphor; the difference is but slight ..." (*Rhetoric* 1406b). Although Aristotle's nuanced view recognizes multiple cognitive principles underlying metaphor (e.g., species to genus, genus to species, species to species, analogy), subsequent thinkers tended to focus almost exclusively on *similarity* as the deep principle underlying these other transfers of meaning. For instance, they reasoned that the basis for applying a term from one species to a different species would be shared similarities between the two species. The principal idea is that metaphor is a means of highlighting literal features shared by both the source and target domains. According to this view, for example, calling Achilles a lion would bring into relief his animal cunning, courage, ferocity, strength, and speed.

There are several brief surveys of the history of philosophical views of metaphor from Aristotle up through the end of the twentieth century (e.g., Ricoeur 1975; Johnson 1981; MacCormac 1985; Cooper 1986). They tend to reveal a mainstream Aristotelian view of metaphor that, with the exception of Nietzsche, is not substantially challenged until the latter half of the twentieth century. The Aristotelian view gets categorized as what Max Black (1954/1955) called the *similarity* or *comparison* theory, according to which a metaphor of the form "A is B" (where "A" is the target domain and "B" is the source domain) is merely a collapsed similarity statement whose meaning is cashed out as "A is similar to B in that both A and B share literal features X, Y, Z, ..." Black realized that this popular default theory makes metaphor for the most part cognitively unimportant, since the meaning of any metaphor could supposedly be spelled out as a list of literal features shared between the source and target domains. Whatever value a metaphor has would thus be based on either its rhetorical punch or its edifying power, but not on any distinctive meaning or truth value, for that meaning must ultimately be expressible as a list of literal similarity statements.

Metaphor as Irreducible Cognitive Process

In the mid-twentieth century, philosophy of mind and language were completely dominated by this comparison view, with its attendant literalist and objectivist assumptions about meaning and truth. Black's (1954/1955) greatest contribution was to shake the whole framework of analytic philosophy that supported the comparison view, by asserting that the cognitive content of a metaphorical expression often could *not* be reduced to some set of literal similarity statements. To account for the cognitive power of a good metaphor, Black proposed the Interaction Theory, which states that metaphors operate through the complex interaction of the systems

of "associated commonplaces" of the source and target domains. Black used the metaphor of metaphor as a "filter" or "screen" through which our understanding of the target domain is organized.

Black then boldly proclaimed that metaphor often plays a distinctive cognitive role in creating new meaning – meaning that cannot be replaced with any set of literal utterances: "It would be more illuminating in some of these cases to say that the metaphor *creates* the similarity than to say that it formulates some similarity antecedently existing" (Black 1954/1955, 285). Black realized the potentially revolutionary metaphysical and epistemological implications of the idea that metaphor is an irreducible cognitive process. The widely-held Comparison (or Similarity) View preserves the traditional literalist idea that the world exists independent of our cognition and that the proper function of literal language is to capture the relevant features of the world. According to such a view, any metaphor would, in principle at least, be completely reducible to literal expressions, and so metaphor would play no crucial role either in how the world is or in how we know it. Consequently, Comparison theories cannot accord metaphor any truly significant role in the way ordinary human beings make sense of their world. Black argued that only some form of Interactionist view would recognize the cognitive irreducibility of metaphor. In a later essay (Black 1977) Black described metaphors as "cognitive instruments" that play a role in the very character of our experience: "some metaphors enable us to see aspects of reality that the metaphor's production helps to constitute. But that is no longer surprising if one believes that the world is necessarily a world *under a certain description* – or a world seen from a certain perspective. Some metaphors can create such a perspective" (454).

In sum, Black played a crucial role in thrusting metaphor into the limelight of philosophical, psychological, and linguistic analysis, even if he was not able to explain fully how metaphor actually works. The idea began to take hold that there might be something cognitively special about metaphor and that this fact might require serious revision of deeply rooted views of thought and language. For the first time in Western philosophy, metaphor had become a necessary, and even a central, topic.

Much of the second half of the twentieth century can be characterized as a development of themes opened up by Black's work and an attempt to answer the kinds of questions a view like his would engender. Those questions were: (1) How do we identify an expression as metaphorical? (2) What is the cognitive mechanism of metaphorical understanding? (3) Does metaphor have a cognitive content (or meaning) that is not reducible to literal concepts or propositions? (4) Where does metaphor fit within a theory of speech acts?

As one would expect, psychologists in the 1960s and 1970s focused chiefly on research questions amenable to their preferred information-processing models of cognition. Consequently, they studied when and how children first learn to interpret expressions as metaphors and later to use metaphors (e.g. Vosniadou and Ortony 1983; Winner 1988). These and other early studies tended to see metaphor comprehension and production as a rather late cognitive development, occurring typically at age seven or eight, but clearly not before age six. However, more recent studies

that have provided more extensive contextual information for metaphor comprehension tasks have suggested that some children as young as 4–5 can understand certain types of metaphors (Maring 2003).

Another experimental issue for psychologists has been whether metaphors take longer to process (as some theories predicted) than do literal expressions. For example, literalists about meaning often assumed that metaphor processing would occur in at least three stages (e.g., Searle 1979). First, the hearer or reader must understand that some expression is not meant to be taken literally. Second, the hearer must then determine possible meanings of the non-literal expression. Third, the precise meaning appropriate to the context of utterance must be selected from among the range of possible meanings. If there were such a three-step process, then it should take longer to understand a metaphorical expression than a literal one. In a recent set of experiments concerning relative comprehension times for various types of literal and metaphorical expressions, Bowdle and Gentner (2005) found that "novel figurative statements took longer to comprehend than conventionalized figurative statements. Indeed, conventional figuratives were comprehended as rapidly as literal comparisons and literal categorizations" (p. 202).

However, a survey of the experimental literature over the past three decades reveals that the issue of comprehension times is highly controversial, and there is probably no settled opinion on this issue at present. The debate often centers on whether or not subjects were given appropriate context for understanding the test expressions, and, indeed, the expressions in the Bowdle and Gentner experiments were given as isolated examples, without any prior context to facilitate their comprehension. Gibbs (1994) had earlier surveyed several psychological studies of this issue and concluded that there is no conclusive evidence that metaphor processing does take longer, once an appropriate interpretive context is supplied to the person trying to understand the expression: "Given sufficient context, people may not need to analyze the literal interpretation of the metaphorical utterances before deriving their intended metaphorical meanings. This was found to be true. When participants read the target sentences at the ends of long contexts, there were no differences in the times to read literal and metaphorical sentences" (Gibbs 1994, 100). Moreover, the psycholinguistic evidence does not indicate that special processing effort is required to understand metaphor:

> The results of the extensive empirical investigations reviewed in this chapter do *not* support many of the hypotheses that figurative language is special and always demands extra work to be interpreted. Instead, the psycholinguistic research indicates that people can understand many instances of figurative expressions effortlessly, without the explicit recognition that such language is special or reflective of deviant thought. (Gibbs 1994, 119)

If psychologists focus chiefly on metaphor processing, philosophers tend to focus primarily on whether metaphors have any special meaning or cognitive content, and consequently on whether they can be bearers of truth (a preoccupation of analytic philosophy of language). Black's Interaction theory spawned a host of accounts of the alleged special cognitive mechanisms of metaphor, including processes such as actualization of connotations previously unnoticed (Beardsley 1962), a Wittgensteinian process of seeing-as or perceiving-as (Aldrich 1968; Hester 1966), a "calculated

category-mistake" or sort-crossing (Goodman 1968), use of a Peircean icon to structure our understanding of the target domain (Henle 1958; Hausman 1989), and suspension of ordinary reference in order to project "new possibilities of redescribing the world" (Ricoeur 1975).

Andrew Ortony's anthology *Metaphor and Thought* (Ortony 1979) brought special attention to the topic by compiling several seminal essays that carried the field forward during the following decade. For example, Searle (1979) attempted to provide a view of metaphor as a distinctive type of speech act governed by a set of rules for determining how an utterance could be recognized as metaphorical and by which its meaning in context could be generated, in the three-step sequence mentioned above. Searle is a literalist about meaning, so his view requires him to state principles by which a speaker could utter a sentence of the form "S is P" (with a literal sentence meaning) and get his hearer to understand that he meant "S is R" (the speaker's utterance meaning). Searle proposed several principles for calculating the appropriate meanings of "R" for a given context in which the speaker utters "S is P," including "that things P are *by definition* R," "that things P are *contingently* R," and "that things P are *generally believed* to be R." However, Searle's fourth principle is most puzzling. It asserts that in some cases none of these first three principles hold, but "nonetheless it is a fact about our sensibility, whether culturally or naturally determined, that we just do perceive a connection, so that *P* is associated in our minds with *R* properties" (Searle 1979, 108). Searle must surely have understood that this "principle" gives no explanation at all, for it amounts only to the assertion that somehow we just do make certain associations. Nevertheless, Searle's statement has some value, insofar as it appears to recognize the role of bodily sensibility in establishing many common metaphors, but it would take several years and dramatic developments in cognitive neuroscience before this statement could be explained.

Challenges to Metaphorical Meaning

Another influential contribution to the Ortony volume was Donald Davidson's deflationary charge that most of the previous work on metaphor was more or less useless, insofar as it rested on the mistaken assumption that metaphor was a semantic phenomenon. Davidson boldly denied that metaphor had any special meaning, beyond the literal meanings of the words used to express it, and he therefore argued that philosophers had been badly misled into regarding metaphors as bearers of truth. His argument was that, since metaphors have no special semantic content or meaning, they do not express propositional truth claims. Instead, Davidson asserted, "we must give up the idea that a metaphor carries a message, that it has a content or meaning (except, of course, its literal meaning)" (Davidson 1978, 45). He claimed that we use this literal meaning to "intimate," or "suggest," or "get someone to notice" something. This intimation is the purpose and power of metaphor, but since "what we notice or see is not, in general, propositional in character" (47),

metaphor is not in the truth-telling business. This relegation of metaphor to the pragmatics of language made Davidson's view instantly popular in some philosophical circles, because it gave analytic philosophers of language a way of preserving their literalist, truth-conditional theories of meaning, while still granting some rationale for use of metaphor (although not to state truths).

Richard Rorty later famously championed Davidson's view by denying (1) that metaphors are semantic, and (2) that they have truth values. Rorty argued that metaphors are irrational ruptures, by which one vocabulary or language-game is supposedly replaced by another. Coming up with a creative new metaphor is a way of establishing a new vocabulary, but not a way of expressing truths within an existing vocabulary. He asserted that "tossing a metaphor into a conversation is like suddenly breaking off the conversation long enough to make a face, or pulling a photograph out of your pocket and displaying it, or pointing at a feature of the surroundings, or slapping your interlocutor's face, or kissing him." (Rorty 1989, 18). In other words, using a metaphor is akin to poking someone with a stick. It may get them to see something, but the poking itself is not a truth-stating gesture. In short, Davidson and Rorty regard the use of a metaphor as a tool for getting another person to notice something, but not by means of any special metaphorical meaning, propositional content, or truth-conditions. Rorty concludes that there is no logic or semantic rationale for the irrational leap to some bold new vocabulary-creating metaphor:

> For genuine novelty can, after all, occur in a world of blind, contingent, mechanical forces ⋯ Analogously, for all we know, or should care, Aristotle's metaphorical use of *ousia*, Saint Paul's metaphorical use of *agape,* and Newton's metaphorical use of *gravitas*, were the results of cosmic rays scrambling the fine structure of some crucial neurons in their respective brains. Or, more plausibly, they were the result of some odd episodes in infancy – some obsessional kinks left in these brains by idiosyncratic traumata. It hardly matters how the trick was done (Rorty 1989, 17).

The furor surrounding Davidson and Rorty on the issue of metaphor stemmed principally from their denial of semantic (and truth-conditional) status for metaphor. To assert that metaphors are not truth-conditional struck at the heart of Black's influential claim that metaphors can present truths that cannot be captured by literal language alone.

Metaphor as Conceptual and Conventional

Because metaphor has traditionally been regarded as a poetic device, or at least as a device of linguistic creativity, it was not regarded as a fundamental principle of ordinary cognition. Even Davidson and Rorty accepted the traditional idea that metaphor is a special occurrence – the proper tool for creativity in language and art. On this received view, metaphor is always taken to be an extraordinary use of language for special purposes, such as in poetry, art, and science. In *Metaphors We Live By* (Lakoff and Johnson 1980), George Lakoff and Mark Johnson challenged this exclusive focus on metaphor as a poetic device for creating new insight,

arguing instead that conceptual metaphor is ubiquitous in our abstract conceptualization and reasoning. Creative uses of metaphor are mostly parasitic on garden-variety metaphorical concepts that structure our ordinary and theoretical understanding alike, in virtually every field of human endeavor, including the sciences (Magnani and Nersessian 2002), mathematics and logic (Lakoff and Nunez 2000), religion (Slingerland 2004), morality (Johnson 1993), law (Winter 2001), politics (Lakoff 1996), philosophy (Lakoff and Johnson 1999), and psychology (Kearns 1987; Fernandez-Duque and Johnson 1999).

Lakoff and Johnson (1980) defined conceptual metaphor as a cross-domain mapping of structure from a source to a target domain, where the two domains are regarded as different in kind. For example, the MOVING TIMES metaphor, found in cultures around the world, recruits the source domain conceptualization of motion through space as a way of understanding the passage of time. In that particular metaphor, times are objects moving toward and then past an observer, the future is in front of the observer, the past is behind, the present is co-located with the observer, and the speed of the moving objects correlates with the "speed" of the passing of time. The metaphor is not "in the words," but rather exists as the conceptual mapping from the source domain of spatial motion onto the target domain of temporal change, as follows.

The Moving Times Metaphor

Source domain (spatial motion)   Target domain (temporal change)
Moving objects   →   Times
Space in front of observer   →   The future
Space behind the observer   →   The past
Relative motion   →   Passing of time

Our linguistic expressions for the passing of time (e.g., Christmas is *coming*, That's all *behind us* now, The hours are *flying by*, Christmas *follows* Thanksgiving) reveal this underlying metaphorical mapping, but those expressions are not the metaphor proper.

The details of the conceptual mapping, plus our knowledge of the source domain, determine the possible inferences we can make about the target domain. In other words, there is a logic of the source domain that, via the conceptual mapping, gets transferred to the inferences we make about the target domain. For instance, in the source domain we know that if two objects A and B are moving toward us in a straight line, with object B following object A, then object B will arrive at our present location after object A. Consequently, in the target domain, we infer that, if Tuesday *follows* Monday, then it will *come after* Monday and *arrive* later.

One of the most distinctive features of the Lakoff and Johnson view is that conceptual metaphors operate mostly beneath the level of conscious awareness to structure our conceptual systems and to guide the inferences we make. As such, they are processed more or less unconsciously and automatically. This claim has subsequently

generated considerable debate in the form of two chief challenges to Conceptual Metaphor Theory: (1) Are conceptual metaphors real? That is, do they actually structure our abstract concepts? What evidence could there be for their reality? (2) If they are real, are they universal across cultures and historical periods, or are they culturally relative?

On the conceptual reality issue, the idea that conceptual metaphors are ubiquitous and structure our most fundamental abstract concepts runs counter to the traditional view that metaphors are chiefly vehicles for creativity or conceptual innovation. The denial of the conceptual reality of metaphor is deep and profound. Most people are default literalists about concepts and meaning, and this carries over even into sophisticated theoretical treatments. Recall that Davidson and Rorty deny metaphor any non-literal semantic content, and therefore do not see them as conceptual. But even those who regard metaphor as a semantic phenomenon often do not see it as essential for our abstract conceptual systems.

This raises the question of what could count as evidence for or against the Conceptual Metaphor hypothesis. John Vervaeke and John Kennedy (1996) argue that the Conceptual Metaphor hypothesis is not empirically falsifiable, because, they claim, whenever an expression is cited as evidence against the existence of a proposed conventional conceptual metaphor, then a defender of the hypothesis will merely identify a new level (another, different metaphor) to explain the new expression: "However, there is nothing in principle to assure that the selection cannot be changed and a new level of generality found to be just as fitting.… The correct level of generality is indeterminate. One immediate consequence of this fact is that, once again, it could be extremely difficult to test the theory of implicit metaphor" (p. 277). Criticisms of this sort have led to discussions of the types of evidence available for conceptual metaphor. Proponents of Conceptual Metaphor Theory have cited at least nine types of evidence, including most prominently:

1. *Polysemy*: Polysemy is the phenomenon of single terms (such as *in, follows, flies*) with multiple related meanings (such as *in* the house, *in* trouble, and *in* the way). Conceptual Metaphor Theory argues that it is the underlying conceptual mappings for the metaphors that explain the relevant meanings of the polysemous terms when they are used in metaphorical expressions (Lakoff and Johnson 1980; Gibbs 1994; Kovecses 2000). For example, spatial senses of *in* (such as, "She is *in* her room") involve a container schema with some entity located within the interior of the container. Via the metaphor STATES ARE LOCATIONS, metaphorical uses of *in* for abstract concepts (e.g., "She is *in* love") appropriate a metaphorical container or bounded location, the state-location of love that one can "fall into" and "fall out of."

2. *Inference Generalizations*: If we really do think using conceptual metaphors, then source-domain inferences should apply for our reasoning about the target domain, too (Lakoff and Johnson 1999). Thus, if an object X is in container A, and container A is within container B, then X is within container B. Therefore, because we can conceive of conceptual categories as metaphorical containers, we reason that, if an object X is *in* category A, and category A is *in* category B, then object X is *in* category B (this is a transitive logical relation).

3. *Novel Extensions*: Lakoff and Turner (1989) give numerous analyses showing how novel poetic metaphors are actually extensions of conventional conceptual metaphors. Poetic treatments of death, they argue, typically rest on ordinary metaphors for death, such as departure from a state-location, change of seasons to winter, or passage to the end of a day.

4. *Semantic Change*: Sweetser (1990) tracked selected cases of semantic change in several Indo-European languages, showing underlying metaphorical patterns for the transformations, such as a MIND AS BODY metaphor. For instance, terms originally used for sensory perception (such as *look, see, grasp, hear*) arise first with their typical bodily/spatial meanings, and then later come to have metaphorical meanings that apply to non-perceptual, non-spatial concepts (such as, "I *see* your point," "That's a slippery concept to *grasp*," and "I *hear* what you're saying").

5. *Spontaneous Gesture*: McNeill (1992, 2005) identified metaphorical gestures in ordinary communication, such as when a speaker holds her hands, palms up, out to her sides and raises first one, then the other, as if testing the weight of two objects, while saying "I couldn't decide whether to go out with Jack last night" (the JUDGING IS WEIGHING metaphor). Such gestures concretely embody the sensory-motor logic of the metaphors they express, and the gestures can have meaning that can complement, or even question, the meaning of the spoken words.

6. *Signed Languages*: Taub (2001) was the first to give an extensive analysis of conceptual metaphor in American Sign Language.

7. *Psychological Experiments*: Gibbs (1994) summarized much of the non-linguistic evidence for conceptual metaphor provided by various kinds of psychological experiments, including priming experiments. As an example of metaphor priming, Boroditsky and Ramscar (2002) subsequently tested the priming for various metaphorical understandings of time, by having subjects experience various types of relative motion prior to their processing of different metaphors based on temporal change as relative spatial motion. There are at least two fundamental spatialization metaphors for time passing – one in which the times move toward the observer, and the other in which the observer moves over a time-landscape – that can be primed by two different experiences of spatial motion, one in which an object moves toward the observer, and the other in which the observer moves over a landscape. The appropriate sensory-motor priming increased the comprehension speed for the appropriate time metaphor.

Determining how to interpret the alleged evidence for conceptual metaphor from these various sources is controversial, since any description of phenomena depends on theoretical assumptions about what counts as data and what forms of explanation are recognized. Those of a literalist orientation will continue to insist that our basic concepts are (must be) literal, so that metaphors can only be alternative ways of expressing what can ultimately be cashed out as literal meaning. However, our inability to provide semantically and inferentially appropriate literal translations for common metaphors has been noted for several decades, even by theorists like Searle (1979) who are espoused literalists themselves. Two alternative literalist views remain popular. One is an Abstractionist position, which claims that whenever

we think we have an implicit conceptual metaphor, there is actually some underlying abstract literal concept that is being expressed via the metaphor. Lakoff and Johnson (1980) argued that the Abstraction theorist would have to be able to explain all of the polysemy and inference generalizations in terms of the alleged generic literal concepts posited. However, the general abstract concepts tend to be so general, so thin, that they lack the necessary semantic and inferential structure that shows up in our language and reasoning. The second alternative, the Homonymy view, treats each instance of a polysemous term as simply independent senses of what just happens to be the same word. The cost for maintaining such a view would seem to be the loss of any ability to explain why a particular term is used for certain concepts, but not for others. There is no adequate Homonymy explanation of why we use the term "in" for "*in* the garden," "*in* the house," "*in* trouble," "*in* time," and "*in* the event that you decide to leave."

The second main controversy, concerning the possible universality of basic metaphors across cultures, stems principally from Lakoff and Johnson's (1980, 1999) claim that conceptual metaphors are typically grounded in aspects of our sensory-motor experience, and so are acquired unconsciously through our bodily engagement with our environment. If this is true, then shared embodiment and shared environments ought to generate universal metaphors for our most basic abstract concepts. Joseph Grady (1997) proposed a theory of "primary" metaphors that are acquired by children simply because of the nature of their bodily experience (in perception and bodily movement) for the kinds of structured environments they inhabit. For example, when an infant or child is held affectionately by a parent, it experiences simultaneously both warmth and a sense of affection and nurturance. This experiential correlation can later become the basis for the conceptual metaphor AFFECTION IS WARMTH, which underlies expressions such as "The audience gave me a *warm* reception," "She was *cold* toward me all evening; in fact, she was a real *block of ice*," and "Our relationship has *cooled* recently." Grady suggests that each of us will normally, and automatically, acquire hundreds of primary metaphors, just by having certain widespread common experiences, such as observing that the moving of an object over a path correlates with a certain passing of time (this would be the basis for the primary MOVING TIMES metaphor). Over the past two decades, a large body of research has appeared on cross-cultural metaphors, much of which has been reported in the journals *Cognitive Linguistics* and *Metaphor and Symbol*. Although there are several good candidates for universal metaphors (such as AFFECTION IS WARMTH, UNDERSTANDING IS SEEING, TEMPORAL CHANGE IS MOTION) based on shared bodily experiences, there is extensive evidence of cultural variation in the elaboration and specification of those metaphors. In Aymara (Nunez and Sweetser 2006) and Mandarin (Ahrens and Huang 2002), for instance, time passing is conceived metaphorically as relative spatial motion, but in both language systems the future is oriented behind the observer and the past is in front – just the opposite of English and virtually all Indo-European languages. Nunez and Sweetser speculate that the past-in-front orientation makes sense, because we can "see"

(that is *know,* via the Knowing Is Seeing primary metaphor) what has already happened (and is in the past). Such differences of past-future orientation do not, however, challenge the bodily grounding or the universality of primary metaphor. What they show, instead, is that cultural variability can exist against a background of shared metaphorical understanding based on universal bodily experiences, such as experiencing the passing of time as one observes an object moving across the landscape. Consequently, although the Temporal Change Is Motion metaphor may be universal, the version of the Moving Times metaphor with the future oriented in front of the observer is clearly not universal. In the end, this must be a question for careful cross-linguistic empirical study, rather than armchair speculation.

Metaphor and Cognitive Neuroscience

The dramatic rise of cognitive neuroscience, with remarkable progress in methods of neural imaging, has recently begun to make it possible to explore neural models of metaphoric competence. These developments have opened up an entirely new way of addressing questions about how we process metaphors. For example, Conceptual Metaphor Theory claims that metaphors are based on the activation of a sensory-motor source domain as a way of conceptualizing and reasoning about some target domain. For this to be correct, it would have to be the case that sensory-motor areas of the brain are activated as we comprehend verbal and nonverbal metaphor-based expressions (as in, "I *see* what you mean" (verbal) or a hand-gesture of extending the arm straight in front of you with the palm facing another person, to communicate that they should stop and come no closer to the speaker (non-verbal)). Consequently, in processing a conceptual metaphor that has a sensory-motor source domain, at least parts of certain areas of the brain responsible for the specific sensory-motor operation in question should be activated when one merely reads or hears a sentence with the appropriate bodily or spatial term upon which the metaphorical understanding is supposedly based. For example, if there really is a conceptual metaphor Understanding Is Seeing, then when you hear "I see your point, but you could shed more light on it," parts of your visual cortex should be activated as you process that expression. For the first time, there is some modest neuro-imaging evidence bearing on this hypothesis. Using fMRI, Tim Rohrer (2001) has shown that both literal and metaphoric sentences using hand terms – such as, "She handed me the apple" (literal) and "He handed me the theory" (metaphorical) – activate primary and secondary hand regions within the primary and secondary sensory-motor cortical maps. Research of this sort is preliminary, but it points the way to future progress on the neural bases of metaphorical cognition, in ways that might one day definitively resolve some of the current debates about the cognitive reality, conceptual structure, and neural mechanisms of metaphor.

Lakoff (2008) has sketched the outlines of a neural theory of metaphor that employs structured or constrained connectionist models. The key idea is to understand the conceptual mapping from source-domain to target-domain in terms of neural mapping – that is, in terms of neural connections between neural maps in different parts of the brain. In Grady's (1997) primary metaphor theory, each metaphor is based on the co-activation of neural ensembles responsible for both the source and target domains; for example, when you experience warmth (source domain) and affection (target domain) at the same time (as the basis of the Affection Is Warmth metaphor), or when you experience the addition of a liquid to a container and observe that the level in the container rises (as the basis of the More Is Up metaphor). In other words, what we call an "experiential correlation" of the source and target domains would be realized in the brain as a "coactivation" of neural clusters. We know that there are topographic and topological maps in various sensory-motor areas of the brain. Each of these neural maps has its own embodied logic, because the firing of certain groups of neurons leads to the firing of other neurons. Lakoff hypothesizes that it is spatial or bodily logics of this sort, preserved in neural maps, that constitute the inference patterns that are used to reason about some abstract domain. Lakoff summarizes the core idea of the neural theory of conceptual metaphor:

> In situations where the source and target domains are both active simultaneously, the two areas of the brain for the source and target domains will both be active. Via the Hebbian principle that *Neurons that fire together wire together*, neural mapping circuits linking the two domains will be learned. Those circuits constitute the metaphor (Lakoff, forthcoming, no page no.).

For the present, the neural work on metaphor comprehension is speculative and preliminary, but as neural imaging technologies and methods continue to develop over the coming decades, we may well see remarkable progress in resolving some of the cognitive issues that currently surround metaphor. Until this research is done, it is not surprising that the traditional view of metaphor continues to hold sway in many disciplines, and in our common folk understanding of language and thought. The reason for this is that the traditional view preserves our deeply-rooted literalist view of concepts and thought. This literalism is built into our conceptual systems and our language for mind, thought, and communication.

The idea that metaphor is a central cognitive process for abstract conceptualization and reasoning calls into question many of the founding assumptions of the traditional view. As we learn much more about brain processes underlying cognition, emotion, symbolic expression, and communicative behavior, we will be in a better position to put to the test previous theories formulated without benefit of good neural models. Productive arguments about the central role of metaphor in human thought are possible today, precisely because so much detailed empirical study is being done within several disciplines and from many methodological perspectives, combining research from linguistics, psychology, philosophy, and neuroscience. This is a complete reversal of the marginalization of metaphor prior to the mid-twentieth century. Metaphor has come of age in the age of cognitive science.

References

Ahrens K, Huang C-R (2002) Time passing is motion. Lang Linguist 3 3:491–519

Aldrich V (1968) Visual metaphor. J Aesthetic Educ 2 1:73–86

Aristotle (1941) Poetics. In: McKeon R (ed) The basic works of Aristotle (trans: Bywater I). Random House, New York

Beardsley M (1962) The metaphorical twist. Philos Phenomenol Res 22(3):293–307

Black M (1954–1955) Metaphor. Proc Aristotelian Soc N.S.55:273–294

Black M (1977) More about metaphor. Dialectica 31(3–4):431–457

Boroditsky L, Ramscar M (2002) The roles of body and mind in abstract thought. Psychol Sci 13(2):185–189

Bowdle B, Gentner D (2005) The career of metaphor. Psychol Rev 112(1):193–216

Cooper D (1986) Metaphor. Basil Blackwell, Oxford

Davidson D (1978) What metaphors mean. Critic Inq 5(1):31–47

Fernandez-Duque D, Johnson M (1999) Attention metaphors: how metaphors guide the cognitive psychology of attention. Cogn Sci 23(1):83–116

Gibbs R (1994) The poetics of mind. Cambridge University Press, Cambridge

Goodman N (1968) The languages of art. Bobbs-Merrill, Indianapolis

Grady J (1997) Foundations of meaning: primary metaphor and primary scenes. Ph.D. Dissertation, Department of Linguistics, University of California, Berkeley

Hausman C (1989) Metaphor and art: interactionism and reference in the verbal and nonverbal arts. Cambridge University Press, Cambridge

Henle P (1958) Metaphor. In: Henle (ed) Language, thought, and culture. University of Michigan Press, Ann Arbor, pp 173–195

Hester M (1966) Metaphor and aspect seeing. J Aesthet Art Critic 25(2):205–212

Johnson M (ed) (1981) Philosophical perspectives on metaphor. University of Minnesota Press, Minneapolis

Johnson M (1993) Moral imagination: implications of cognitive science for ethics. University of Chicago Press, Chicago, IL

Kearns M (1987) Metaphors of mind: in fiction and psychology. University of Kentucky Press, Lexington

Kovecses Z (2000) Metaphor and emotion: language, culture, and body in human feeling. Cambridge University Press, Cambridge

Lakoff G (1996) Moral politics: what conservatives know that liberals don't. University of Chicago Press, Chicago

Lakoff G (2008) The neural theory of metaphor. In: Gibbs R (ed) The Cambridge handbook of metaphor and thought. Cambridge University Press, Cambridge

Lakoff G, Johnson M (1980) Metaphors we live by. University of Chicago Press, Chicago, IL

Lakoff G, Johnson M (1999) Philosophy in the flesh: the embodied mind and its challenge to western thought. Basic Books, New York

Lakoff G, Nunez R (2000) Where mathematics come from: how the embodied mind brings mathematics into being. Basic Books, New York

Lakoff G, Turner M (1989) More than cool reason: a field guide to poetic metaphor. University of Chicago Press, Chicago, IL

MacCormac E (1985) A cognitive theory of metaphor. MIT Press, Cambridge, MA

Magnani L, Nersesian N (eds) (2002) Model-based reasoning: science, technology, values. Kluwer, New York

Maring B (2003) The metaphorical bases of children's developing theories of mind. Unpub. Ph.D. Dissertation, Department of Psychology, University of Oregon

McNeill D (1992) Hand and mind: what gestures reveal about thought. University of Chicago Press, Chicago, IL

McNeill D (2005) Gesture & thought. University of Chicago Press, Chicago, IL

Nunez R, Sweetser E (2006) With the future behind them: convergent evidence from aymara language and gesture in the cross-linguistic comparison of spatial construals of time. Cogn Sci 30(3):401–450

Ortony A (ed) (1979) Metaphor and thought. Cambridge University Press, Cambridge

Ricoeur P (1975) The rule of metaphor: multidisciplinary studies of the creation of meaning in language (trans: Czerny R with McLaughlin K, Costello J). University of Toronto Press, Toronto

Rohrer T (2001) Understanding through the body: fMRI and ERP studies of metaphoric and literal language. Paper presented at the 7th International Cognitive Linguistics Association conference. University of California at Santa Barbara, Santa Barbara, CA, July 2001

Rorty R (1989) The contingency of language. In Contingency, irony, and solidarity. Cambridge University Press, Cambridge, pp 3–22

Searle J (1979) Metaphor. In expression and meaning. Cambridge University Press, Cambridge, pp 76–116

Slingerland E (2004) Conceptual metaphor theory as methodology for comparative religion. J Am Acad Religion 72(1):1–31

Sweetser E (1990) From etymology to pragmatics: metaphorical and cultural aspects of semantic structure. Cambridge University Press, Cambridge

Taub S (2001) Language from the body: iconicity and metaphor in American sign language. Cambridge University Press, Cambridge

Vervaeke J, John Kennedy J (1996) Metaphors in language and thought: falsification and multiple meanings. Metaphor Symbol Activity 11(4):273–284

Vosniadou S, Ortony A (1983) The emergence of the literal-metaphorical-anomalous distinction in young children. Child Dev 54:154–161

Winner E (1988) The point of words: children's understanding of metaphor and irony. Harvard University Press, Cambridge, MA

Winter S (2001) A clearing in the forest: law, life, and mind. University of Chicago Press, Chicago, IL

Phenomenology and Cognitive Linguistics

Jordan Zlatev

Introduction

The purpose of this chapter is to describe some similarities, as well as differences, between theoretical proposals emanating from the tradition of phenomenology and the currently popular approach to language and cognition known as *cognitive linguistics* (hence CL). This is a rather demanding and potentially controversial topic. For one thing, neither CL nor phenomenology constitute monolithic theories, and are actually rife with internal controversies. This forces me to make certain "schematizations", since it is impossible to deal with the complexity of these debates in the space here allotted.

Phenomenology was founded by Edmund Husserl over a century ago as a new foundational philosophy, based on "the careful description of what appears to consciousness precisely in the manner of its appearing" (Moran 2005: 1). While students and descendants such as Heidegger, Merleau-Ponty, Sartre and Gurwitsch, to mention only the most famous ones, developed this tradition in their own directions, I believe that the figure of Husserl arches over these, and in the same way that he urged to "go back to the things themselves", it is necessary to "go back to Husserl" in order to truly understand what phenomenology is about. I think also that it is fair to say that the past decade has brought about a reappraisal of Husserl, along with the increased familiarity with his late and mainly posthumously published work. Husserl's high relevance for present "hot topics" such as the nature of consciousness, the embodiment of subjectivity and the intersubjective constitution of the lifeworld, is widely acknowledged (Thompson 2001; Zahavi 2003; Depraz 2001; Moran 2005; Gallagher 2005). Hence, I will mostly regard Husserl as "metonymic" for phenomenology, and make brief excursions into the work of other phenomenologists, including moderns ones, only when this is called for. Furthermore, other chapters in this *Handbook* are bound to treat the debates within phenomenology in much more detail.

J. Zlatev
Centre for Cognitive Semiotics, SOL, Lund University, 221 00 Lund, Sweden

CL emerged in the late 1970s, to a large extent as a reaction to the dominance of formalist approaches to language and cognition such as *generative linguistics* (Chomsky 1975), *logical semantics* (e.g. Montague 1974), and *computationalist cognitive science* (Gardner 1987). The "founding fathers" of CL (cf. Geeraerts and Cuyckens 2007b) are typically considered to be George Lakoff, Ronald Langacker and Leonard Talmy, all California-based linguists. However, it was clear from the start that the ambitions of at least some CL researchers involved not just a new linguistics, but a novel philosophical theory labeled "experientialism" or "embodied realism" (Lakoff and Johnson 1980, 1999; Lakoff 1987; Johnson 1987). Based on professed "empirical discoveries" such as the mind's embodiment, the (predominantly) unconscious nature of thought and the metaphorical nature of abstract concepts, Lakoff and Johnson (1999: 1) claim that their approach constitutes a revolution in Western thought:

> What would happen if we started with these empirical discoveries about the nature of the mind and constructed philosophy anew? The answer is that empirically responsible philosophy would require our culture to abandon some of its deepest theoretical assumptions.

It is in this quest for providing a new philosophy – grounded in "experience", and with a central role played by the living body – that one may be tempted to find the clearest rapprochement between CL and phenomenology. The insistence of an "empirically responsible philosophy" is furthermore reminiscent of present discussions of "naturalizing" phenomenology (cf. Gallagher and Broested-Soerensen 2006). However, a more careful consideration of the writings of Lakoff and Johnson shows that their version of "experientialism" is quite inconsistent with phenomenology, which I will attempt to show in this chapter.

But while these authors are commonly taken as representing "the" philosophy of CL, this is not the case. Over the past 10–15 years, CL has considerably diversified and many have formulated alternative positions on crucial issues such as the nature of meaning, representation, grammar – and, especially relevant for this chapter – consciousness, embodiment and intersubjectivity. Zlatev (1997, 2007) has criticized the Lakoff-Johnson approach for lacking a socio-cultural perspective and, in agreement with Itkonen (2003), for failing to account for the essential *normativity* of language (see Section on Methodology: Phenomenological and "Empirical" Methods). Haser (2005) has presented an extended and influential critique of the philosophy of Lakoff and Johnson, with emphasis on their theory of "conceptual metaphor", showing that despite occasional claims to the contrary linguistic meaning is regarded by "mainstream CL" as a private, mental, "in the head" phenomenon. This is at odds with the meaning-externalism that has dominated philosophy since Wittgenstein (1953), and I will argue that at least one form of externalism is consistent with a phenomenology acknowledging the distinction between linguistic *intuition* and *introspection* (cf. Section Methodology: Phenomenological and "Empirical" Methods). Two recent anthologies within the main publication series of CL, "Cognitive Linguistic Research" edited by Hampe (2005) and Ziemke et al. (2007), have demonstrated the amount of theoretical divergence with respect to, respectively, the central theoretical notion of *image schemas* (cf. Section on Image Schemas) and the meta-theoretical concept of *embodiment* (cf. Section on Embodiment).

The most authoritative summary of CL research up to date, *The Oxford Handbook of Cognitive Linguistics* (Geeraerts and Cuyckens 2007a), represents much of the diversity mentioned in the previous paragraph, including the introspection-based analyses of Talmy (2000), the most developed theory of grammar within CL, the *Cognitive Grammar* of Langacker (1987, 2006), and the influential theory of *Conceptual Integration* ("blending") (Fauconnier and Turner 2002; Turner 2007), with roots in literature studies. Interestingly, the editors (Geeraerts and Cuyckens 2007b) point out in their introductory chapter to a rising tension within CL, which they formulate in epistemological/methodological terms:

> If we understand *empirical methods* to refer to forms of research (like corpus linguistics, experimentation, and neurological modeling) that do not rely on introspection and intuition but that try to ground linguistic analysis on the firm basis of objective observation, then we can certainly witness a growing appeal to such empirical methods within Cognitive Linguistics … (ibid: 16) /…/ Because meanings do not present themselves directly in the corpus data, will introspection not always be used in any cognitive analysis of language? (For an explicit defense of such a position, albeit in terms of "intuition" rather than "introspection" see Itkonen 2003). (ibid: 18)

One of the major contentions of this chapter is that a phenomenological perspective would allow the resolution of this tension. In brief: in a similar way that phenomenological psychology can be argued to serve as prerequisite for empirical psychology,[1] a *phenomenological (cognitive) linguistics* sets the conditions for linguistic "empirical methods" (in the narrow sense of the term used by Geeraerts and Cuyckens, which is quite common nowadays). Thus, while the first part of my presentation will be mostly critical, the motivation is not negative per se, but to clear the ground for a *true* rapprochement between CL and phenomenology. Since many of the controversies on which I will touch concern the "naturalization" of CL, these debates may also have significance for corresponding debates concerning phenomenology itself (e.g. Varela 1996; Sheets-Johnstone 2004; Overgaard 2004).

Fundamental Issues: "Experientialism" Versus Phenomenology

Lakoff and Johnson provide occasional references to "phenomenology", but in ways that are quite confusing, and sometimes even contradictory. For instance, Lakoff and Johnson (1980: 181–182) pay homage to "the central insights of the

---

[1]"The experimental method is indispensible. /…/ But this does not alter the fact that it presupposes what no experiment can accomplish, namely the analysis of conscious life itself. /…/ Phenomenological psychology is interested primarily in the necessary a priori of every possible empirical psychology" (Kockelmans 1967: 425, 447, quoted by Itkonen 2003: 110). The complementary relationship between phenomenological psychology (Husserl [1925] 1977; Gurwitsch 1964) and phenomenological philosophy, focusing respectively on the ego and world poles of the intentional relation is discussed by Sonesson (2007).

phenomenological tradition, such as the rejection of epistemological foundationalism, the stress on the centrality of the body in the structuring of experience, and the importance of that structure in understanding." At the same time, only a few pages later they refer to "the tradition of Husserl" as one form of the criticized "objectivist tradition of Western philosophy" (ibid: 195) – an apparent self-contradiction, also noticed by Haser (2005: 88). In what is probably the most philosophically elaborated work on experientialism up to date, *The Body in the Mind*, Johnson (1987: xxxvii) writes: "some of my most important claims are anticipated in the work of philosophers who might claim allegiance to phenomenology of the post-Husserlian varieties" and describes his method as "a form of descriptive or empirical phenomenology". At the same time, the name of the founder of phenomenology, Husserl, always seems to appear in a negative context in experientialist writings, and that of Merleau-Ponty, whose work is possibly most relevant for the proposed "embodied realism", hardly at all (cf. Nerlich and Clarke 2007).

In this section I will focus on what may be called the "official philosophy" of CL, *experientialism*, and point out (mostly) inconsistencies with phenomenology. At the same time, I will introduce some of the criticism directed at Lakoff and Johnson form less-central representatives of CL, and (conversely) show their affinity with a phenomenological perspective. One of these, Esa Itkonen (1978, 1997, 2003), has been especially influential over CL during recent years (cf. the quotation by Geeraerts and Cuyckens 2007b in the previous section; also Zlatev 2007; Harder 2007).

Metaphysics

Experientialism is contrasted with "Objectivism" on the one hand, and with "Subjectivism" on the other (Lakoff and Johnson 1980: Chapters 25–29). The first is characterized as an "offending cluster of assumptions that has lead to this blindness towards imagination" (Johnson 1987: ix–x), one of these assumptions being formulated as: "there is an objective reality, and we can say things that are objectively, absolutely and unconditionally true about it" (ibid: 187). In some contexts, Objectivism is called a "myth" (along with Subjectivism and Experientialism), apparently meaning some sort of pre-theoretical, culture-specific conceptual framework. In others, however, it is explicitly stated that "objectivist metaphysics … is a false theory" (Lakoff 1988: 124). As Haser (2005: Chapter 5) shows in detail, it is indeed hard to understand what exactly "objectivism" refers to, and why so many diverse philosophers from Frege and "the followers of Husserl" to Davidson and Goodman are claimed to fall prey to its errors. "Subjectivism" – represented by "contemporary interpretations (probably *mis*interpretations) of recent Continental philosophy, especially the traditions of phenomenology and existentialism" (Lakoff and Johnson 1980: 223) – is presented as the anti-thesis. It is characterized by statements of "café phenomenology" (ibid: 224) such as: "Meaning is always a matter of what is meaningful and significant to a person. What an individual finds

meaningful and is significant to him are matters of intuition, imagination, feeling and individual experience" (ibid: 224).

Experientialism is presented as a "synthesis", agreeing with Objectivism on the existence of an "external reality", but with Subjectivism that this reality is mediated by human experience. A correspondence theory of truth is possible only from "within" a given conceptual scheme. Despite claims to the contrary, the doctrine of experientialism thus implies not only what Husserl called "anthropologism" (Husserl 2001 [1900], Prol #36), but cultural relativism: "Being objective is always relative to a conceptual system and a set of cultural values." (Lakoff and Johnson 1980: 227). In general, Lakoff and Johnson avoid paying explicit tribute to other philosophers, but both Lakoff (1987) and Johnson (1987) acknowledge Putnam's (1981) *internal realism* as making essentially the same claims, and indeed Lakoff (1987) refers to "internal realism" as being a valid theoretical extension of commonsense "basic realism", while Objectivism is an invalid one. More recently, Johnson and Rohrer (2007: 21) refer to their doctrine as "embodied realism", and see its predecessors in American Pragmatism, quoting especially approvingly Dewey (1981 [1925]: 198): "to see the organism in nature, the nervous system in the organism, the brain in the nervous system, the cortex in the brain is the answer to the problems that haunt philosophy".

The foundations of phenomenology are quite different. For Husserl (and "his followers") what has primacy is not "objective reality", "consciousness", "inter-action" or biology, but the Lifeworld (*Lebenswelt*), as summarized by Moran (2005: 9): "As conscious beings, we always inhabit – in a pre-theoretical manner – an experiential world, given in advance (*vorgegeben*), on hand (*vorhanden*), and always experienced as a unity". This world is a correlate of a pervasive and universal "naïve" attitude, *the natural attitude* in which we focus our attention on what is "out there" – but not on ourselves as experiencers or on the act of experiencing. Various forms of objectivism such as metaphysical realism ("things in themselves", "the world as such") arise according to Husserl as "abstractions" or "absolutizations" of the natural attitude: "the objectivist, scientific attitude and the formal mathematical attitude are both abstractions from the natural attitude and in a sense presuppose it" (Moran 2005: 55). The point is thus not that such absolutizations are culturally-relative myths or even less "false theories", but that they have a tendency to be, so to say, doubly blind: both that they presuppose the natural attitude and that though basic, the natural attitude is only one among a multitude of attitudes.

Husserl recognized namely that the Lifeworld allows for *a plurality of attitudes*, which again are not "cultural schemes" but possible experiential relations that all human beings can stand in with respect to the world, actualized by particular circumstances. For example, a special type of absolutization of the natural attitude that is particularly prominent (and pernicious) is the modern "technological calculating attitude" (Moran 2005: 237). At the same time, such absolutizations make some of the alternative attitudes appear clearly, often as reactions to the distortions of the Lifeworld that the former introduce. One of these is the *psychological attitude,* in which we thematize our "inner" mental life, rather than "outer" world. However,

when these themselves are absolutized in philosophical doctrines, such as the subjective idealism of Berkeley, they too become no less distorting of the Lifeworld.

However, two attitudes have privileged status for Husserl. The first is *the personalistic attitude*. At least in his later work Husserl claimed "that the natural attitude is actually reached through a self-forgetting of the self or ego of the personalistic attitude, through abstraction from the personal, which presents the world in some kind of absolutized way, as the world of nature" (Moran 2005: 216). Importantly, the personalistic attitude reveals others as *persons* (rather than objects) on a par with the self and in this way makes clear "the intersubjective constitution of the world" (Husserl 1970: #49). The second is the *theoretical* or *philosophical attitude*, developed in the West first in Ancient Greece, allowing for the un-concealment of the natural attitude as such: "The natural attitude is self-concealing; the theoretical attitude, on the other hand, is self-involving and self-aware and ultimately universal, transforming all human praxis (Hua 6: 334)" (quoted by Moran 2005: 150). Phenomenology, including the methods of *epoché* and the reduction, is nothing but a "purification" of this attitude: "this exploration of the dimensions of givenness of objectivity to subjectivity" (Moran 2005: 106).

In this brief exposition, I have deliberately skipped over Husserl's move to transcendental idealism and "absolute consciousness" in *Ideen I*, which made many of his former students part company with him, and which is admittedly rather hard to swallow for most modern thinkers.[2] Still, even if we take that into account, neither Husserl nor any of the other phenomenologists, including Sartre and his insistence on a "pre-positional" *cogito* (Sartre 1956 [1943]) may be properly characterized as either "objectivist" or "subjectivist": all varieties of phenomenology are fervent critics of both.

Furthermore, from the standpoint of Husserl, Gurwitsch (1964), and arguably also Merleau-Ponty (see below), the metaphysics of "experientialism" is yet another distortion of the Lifeworld, and not a particularly coherent one at that. The culture-relativism (objectivity being "always relative to a conceptual system and a set of cultural values") is perhaps consistent with the later work of Heidegger, but for Husserl, this is self-refuting: why accept Lakoff and Johnson's "experientialism", combining what is considered valid (on what grounds?) from objectivism and subjectivism, as anything more than a culture-relative "myth"? The emphasis on pan-human structures of "embodiment" can be seen as a step to address this objection. However, even if that move were successful, which I argue in section on Intersubjectivity is not the case, that would make experientialism species-relative, and like Kant's "anthropologism" ultimately unsatisfactory. The major problem from a phenomenological perspective, however, is that "experientialism" is based on neither of the two most authentic attitudes toward the Lifeworld – the personalistic and the philosophical – but oscillates uneasily

_____

[2] However, some modern philosophers of mind, bedeviled with the failures of both physicalist and dualist theories to explain consciousness, appear to be willing to adopt a form of monism which is reminiscent of Husserl's transcentendal idealism, e.g. Honderich (2006) theory of "radical externalism" or "consciousness as existence".

between the psychological/subjective and the *biological* attitudes. When it takes the first option – made most explicit in the writings of one of the other two "founding fathers", Talmy (see section on Methodology: Phenomenological and "Empirical" Methods) – and e.g. describes meanings as mental entities, it faces the problems of subjectivism. Since this is clearly unsatisfactory, there is a predictable move in the opposite direction, with claims of reducing meaning and experience to neurobiology, as evidenced in the most recent work of Lakoff (Gallese and Lakoff 2005; Dodge and Lakoff 2005). Consider statements such as the following: "Linguistic structure is below the level of consciousness because the brain structures that compute them are unconscious" (ibid: 86) and "The brain is thus the seat of explanation for cognitive linguistic results" (ibid: 70).

From a phenomenological perspective, this is clearly an objectivist absolutization of the natural attitude, which is ultimately reductionist.[3] As pointed out by (Husserl 1989 [1952], #51: 201): "He who sees everywhere only nature, nature in the sense of, as it were, through the eyes of natural science, is precisely blind to the spiritual sphere, the special domain of the human sciences". Even the most "naturalized" of the classical phenomenologists, Merleau-Ponty, argued rather for "abandoning the body as an object … and going back to the body which I experience at this moment" (Merleau-Ponty 1962: 75).

Methodology: Phenomenological and "Empirical" Methods

In a recent "study in phenomenological semiotics" Sonesson (2007: 90) has argued that linguistics as a whole presupposes an explicit or implicit phenomenological

---

[3] Some (like an anonymous reviewer) are liable to dispute this claim, and point out that Lakoff and Johnson (1999) postulate different "levels of embodiment" (cf. Section on Embodiment), while Rohrer (2007a,b) explicitly argues that his "levels of investigation" framework in not reductionist: "research in embodied cognitive science should not seek to reduce such phenomena to another level but should instead bridge across these levels" (Rohrer 2007a: 346). Lakoff and Johnson (1999) describe their ontological position as being one of "noneliminative physicalism" (ibid: 109), where "each level is taken as real, as having a theoretical ontology necessary to explain phenomena. … explanation and motivation flow in both directions." (ibid: 113). However, while this may qualify as an *epistemological* non-reductionism, ontologically Lakoff and Johnson are physicalists, accepting without any argument "the lack of any mind-body gap" (ibid: 96). Also, just considering that 4 of the 6 "levels of investigation" in Rohrer's "non-reductive" framework deal with increasingly high-grained analysis of the brain ("Neural systems", "Neuroanatomy", "Neurocellular systems" and "Subcellular systems"), while the two highest: "Communicative and cultural systems" and "Performance domain" are characterized as "Multiple central nervous systems" and "Central nervous systems" shows what is *really real* for this strand of cognitive linguistic thinking. A non-reductionist neuroscientific cognitive linguistics is indeed possible, but as in neurophenomenology (Varela 1996), that would mean not privileging the objective, third-person perspective, but rather starting from, and keeping a focus on, the *experiences* of speakers, while looking for correlations with these in e.g. neuroimaging studies. I am not aware that any CL-researchers have carried out such projects, but they are of course quite possible.

method, i.e. a careful analysis of what appears in consciousness when we reflect on our knowledge and use of language:

> All human and social sciences which aspire to discover regularities, such as linguistics and other semiotic sciences, *necessarily start out from phenomenology* – and we should be happy if those phenomenological investigations sometime manage to be as meticulous as those of Husserl and Gurwitsch. (emphasis added)

Johnson also claims to be employing a method of "descriptive or empirical phenomenology" (Johnson 1987: xxxvii) or "informal phenomenological analysis of the structural dimensions of sensory-motor experience" (Johnson 2005: 21). In practice, Johnson includes references to everyday situations in which certain invariants of experience are focused on and labeled "image schemas" (for more discussion of this notion, cf. Section on Image Schemas). Since the linguistic expressions used to refer to these are also used in other, and more abstract, contexts the latter are postulated to be derived from "conceptual metaphors". For example, the following passage is meant to show the prevalence of the CONTAINER image schema:

> You wake *out* of a deep sleep and peer *out* from beneath the covers *into* your room. You gradually emerge *out* of your stupor, pull yourself *out* from under the covers, climb *into* your robe, stretch *out* your limbs, and walk *in* a daze *out* of the bedroom and *into* the bathroom. You look *in* the mirror and see your face staring *out* at you. … Once you are more awake you may even get lost *in* the newspaper, might enter *into* a conversation, which leads to your speaking *out* on some topic. (Johnson 1987: 31, emphasis in original)

Walking-into-the-bathroom is presumably an instance of the "pre-conceptual" schema derived from sensory-motor experience, while getting-lost-in-the-newspaper does not have the CONTAINER structure in itself, but rather inherits it via a "conceptual metaphor" (see Johnson this volume). However, while walking into bathrooms and being lost in newspapers are clearly *experiences*, and as such can be subjected to phenomenological analysis, the structures claimed to underlie them are theoretical constructs which may not. Hence, it is unsurprising when Johnson claims:

> However, we must keep in mind that phenomenological analysis is *never enough*, because image schemas typically operate beneath the level of conscious awareness. That is why we *must go beyond phenomenology* to employ standard explanatory methods of linguistics, psychology and neuroscience. (Johnson 2005: 21, emphasis added).

This seems rather contradictory – first phenomenological analysis (even if "informal") is used to uncover structures, which are later claimed to be accessible only through the "standard explanatory methods" of natural science, since they operate "beneath the level of conscious awareness". The last quotation is also revealing of a highly problematic conception of both phenomenology and linguistics. First, phenomenology is mostly concerned not with structures and processes which "typically operate" on/ above (?) "the level of conscious awareness", but rather the contrary: with what is normally non-thematic, or in the "margins" of consciousness, but which through reflection and phenomenological analysis can be brought into awareness. Sonesson (2007) gives the following simple example: while speakers are not aware of the

inventory of phonemes in their language, by conducting a linguistic analysis finding "minimal pairs": e.g. *roar* vs. *lore*, /r/ and /l/ can be identified as phonemes in English, but not in a language such as Japanese.

Secondly, by placing it on a par with psychology (presumably "empirical") and even more so "neuroscience", Johnson is aligning linguistics with the natural sciences. Itkonen (1978, 2003, 2005, 2008) argues against such a misguided application of a "physicalistic attitude":

> For years now, leading representatives of theoretical linguistics have been arguing that humans, being guided by a blind "language instinct", can be described in physico-biological terms. … this conception has been shown to be fundamentally false. Humans are also, and crucially, social, normative, and conscious beings. (Itkonen 2003: 151)

To see why Johnson's reasoning is inconsistent with phenomenology – while the claims of Itkonen in fact are – let us look again at Johnson's passage illustrating the "image schema" CONTAINER given earlier. Two obvious features (bypassed by the author, as well as by the extensive CL literature on "image schemas") are (a) the repeated use of the second-person pronouns *you* and *your*, and (b) the fact that while the passage is supposed to illustrate *private*, and even unconscious, "sensory-motor experience" it is being mediated by a *public* language, in particular, the emphasized terms *in(to)* and *out of*. If the structure was truly *sensory-motor* by nature, Johnson would be at most justified to say "I" and "mine", rather than "you" and "your".[4] The reason for this is that "experientialism" lacks a principled account of *intersubjectivity* (see section on Intersubjectivity), despite occasional claims that what are being described are "shared meanings". The meanings of the expressions illustrating these putative underlying structures, on the other hand, are intersubjectively shared between Johnson and his readers, on the basis of their common knowledge of English, along with the relevant cultural background involving robes, mirrors, newspapers etc. Following Wittgenstein (1953), thus, we may question the legitimacy of trying to explain what is "out in the open" (the *conceptual*) with "hidden" *causal* structures and processes. Of course, when investigating truly causal processes which are by their nature inaccessible to consciousness, one *should* "go beyond phenomenology" and employ "empirical methods". Such investigations are, however, qualitatively different and, as argued at length by both Husserl and Wittgenstein, methodologically secondary. Itkonen makes this point as follows:

> Knowledge of the rules of language precedes any experiment. Imagine an experiment that would show that *cat* really means 'dog', or that in English the definite article does not precede but follows the noun. We would refuse to accept the results of these experiments because *in this domain our common (and pre-experimental) knowledge defines how things are*. It would be an analogous mistake that the notion of e.g., a centimeter is the result, and not a precondition, of experiments where length is measured. (Itkonen 2003: 137, emphasis added)

---

[4]And even less the first-person plural "we", as done by Rohrer (2007b: 35) in paraphrasing the same passage.

The domain Itkonen refers to coincides with those aspects of the Lifeworld which consist of what Searle (1995) calls "institutional facts" rather than "brute facts".[5] It is a phenomenological fact that "the rules of language" (including the meanings of words) are known on a *pre-theoretical level*, as shown by the universal existence of linguistic *intuitions* of correctness (e.g. *John loves Mary*) or incorrectness (*Loves John Mary), even in the most "naïve" speakers (cf. Zlatev 2008). The job of the linguist or philosopher is first of all to provide a *theoretical* explication (rather than causal explanation) of these rules, on the basis of such intuitions. Langacker's explications in terms of pictorial diagrams, Montague's in terms of higher-order logics, or Chomsky's in terms of increasingly complex tree-diagrams thus have, whether their authors are aware of this or not, the same ontological status and methodological basis. *Intuition* is understood by Itkonen as defined by Cohen (1986: 75): "the immediate, unreflective inclination, without argument or inference, to judge that p (and that anyone who faces the same issue ought also to judge that p), where the judgment that p is a kind that is in principle not checkable by sensory experience or by accepted methods of calculation" and this is consistent with Husserl's notion of *categorical intuition* (Husserl 2001 [1900]).

Intuition has an irreducibly *normative element* (note the "ought to" in Cohen's definition) and hence unlike *introspection* is not something private, and "subjective", even though it consists of acts of consciousness. Therefore Itkonen (1978, 2003) has correctly, though controversially, asserted that linguistics is not (primarily) an "empirical science" based on the observation of spatio-temporal occurrences of events, but a "normative science" studying intersubjectively binding norms, (primarily) through intuition. The term "normative science", however, may be confusing, since it usually refers to *prescriptive* fields such as rhetoric, while linguistics, like phenomenology is basically *descriptive*, though what it studies are norms (rules, conventions) rather than concrete entities such as utterances.[6]

The failure to distinguish between intuition and introspection has lead to serious misunderstandings of the methods of both phenomenology (cf. Thomasson 2002, including the quote from Dennett (1987: 154) claiming the phenomenology is based on "some sort of introspection") and (cognitive) linguistics. Talmy (2000: 4), for example, makes this double error in a single passage that has been often quoted as representative of the "non-empirical" camp within CL:

---

[5] Though, of course, Husserl would never have accepted the existence of "brute facts", and rather have said that they belong to another part of the Lifeworld than the institutional ones. A criticism that may be leveled at the early Husserl is that he, similar to Gibson (1979) hardly recognized the existence of the latter (cf. Sonesson 1989).

*By convention, linguists prefix "ungrammatical sentences"

[6] Again, depending on one's purposes one may need to "go beyond" such description and seek "explanations" in terms of evolution, ontogenetic development, history, neuroscience etc. Hence, there are subfields of linguistics devoted specifically to such investigations in which phenomenology is indeed "never enough": psycholinguistics, neurolinguistics, diachronic (historical) linguistics etc. The grounds for these, including notions such as "sentence", "word", "sense" are, however, set by (informal) phenomenological methods.

> For cognitive semantics, the main object of study itself is qualitative mental phenomena as they exist in awareness. Cognitive semantics is thus a branch of phenomenology /…/ As matters stand, the only instrumentality that can access the phenomenological content and structure of consciousness is that of introspection.

The problem is that this would appear to make phenomenological methods subjective and unreliable, and therefore "empirical methods" including those of psycholinguistics, neurolinguistics and the statistical analysis of multi-million word computerized texts (corpora) are currently seen by an increasing number of cognitive linguists as being *necessary* for providing "a truly scientific" theory of meaning and language.

This conflation of introspection and intuition (and phenomenology) is unfortunately pervasive within CL, and it is (I would hold) one of the major reasons for the tension between the "psychologistic" and "physicalistic" attitudes within CL noted earlier. On the one hand, language, and especially meaning is usually treated as something subjective and psychological: "qualitative mental phenomena" (Talmy 2000), "the mental image associated with your basic-level concept …" (Lakoff 1987: 129), "mental scanning" and "mental simulation" (Langacker 1987). Harder (2007: 1247) states this clearly: "Cognitive Linguistics focuses on mental, conceptual entities as legitimate objects of description in their own right". To counter the objection that this would imply subjectivism and anything-goes relativism, the mental is also said to be identical with, or explained by the bio-physical: "an ordered conception necessarily incorporates the *sequenced occurrence* of cognitive events as one facet of its neurological implementation" (Langacker 1986: 455) and (the previously quoted) "The brain is thus the seat of explanation for cognitive linguistic results" (Dodge and Lakoff 2005: 71).

Thus, we are faced with something of a paradox: starting from a kind of naïve phenomenology and an apotheosis of "experience" and "imagination", some of the most prominent cognitive linguists have been led to positions that may be called anti-phenomenological (cf. Section on Embodiment). The resolution of this paradox, defended by some current researchers more or less closely related to CL, is that linguistic meaning and structure are neither subjective (in the mind), nor objective (in the brain), nor in the strange modern conception of the "mind/brain" (cf. Sinha 1999), but are fundamentally *intersubjective* (Itkonen 2003, 2008; Haser 2005; Zlatev 2005, 2007, 2008; Sinha and Rodríguez 2008; Harder 2007). As such they are properly studied through phenomenological methods such as intuition, empathy and explication. "Empirical methods" are also important, but they come in second place.

Intersubjectivity

Towards the end of his monograph, Johnson (1987: 176) summarizes his discussion of "image schemata and metaphorical projections" as follows: "These embodied and imaginative structures of meaning have been shown to be shared, public and "objective",

in an appropriate sense of objectivity". However, it not clear where and how this has "been shown". The use of the second-person pronouns, commented in the previous sub-section, and even the first-person plural: "An image schema is a recurring dynamic pattern of *our* perceptual interactions and motor programs that gives coherence and structure to *our* experience …" (Johnson 1987: xiv, my emphasis) is hardly justified. The closest is probably a neo-Kantian conception that without such "structures of imagination" experience would be chaotic, combined with an empiricist/pragmatist view of "recurrent … interactions" with the environment, and finally a naturalistic conception of "embodiment": "commonalities […] exist in the way our minds are embodied" (Lakoff and Johnson 1999: 4) and "we all have pretty much the same embodied basic-level and spatial-relations concepts" (Lakoff and Johnson 1999: 107).

It is Itkonen (2003, 2008), again, who from phenomenological and Wittgensteinian positions has exposed the fact that experientialism has "no conceptual apparatus" (Itkonen 2003: 148) to account for "public shared meaning" (Johnson 1987: 190), despite the best intentions. Both Johnson (1987: xxx–xxxi) and Itkonen (2003: 145–146) quote a well-known passage from Frege (1966 [1892]) (though in different translations):

> The reference and sense [*Sinn*] of a sign are to be distinguished from the associated idea. If the reference of the sign is an object perceivable by the senses, my image of it is an internal image, arising from memories of sense impressions which I have had, and acts, both internal and external, which I have performed. Such an idea is often saturated with feeling; the clarity of its separate parts varies and oscillates. The same sense is not always connected, even in the same man, with the same idea. The idea is subjective: one man's idea is not that of another. … This constitutes an essential distinction between the idea and the sign's sense, which may be *the common property of many* and therefore is not a part or mode of the individual mind. (emphasis added)

The interpretations that Johnson and Itkonen make of *Sinn*, however, are completely at odds. Johnson writes:

> Frege thought he needed this strange third realm [i.e. *Sinn*, "sense"] to ensure the objectivity of meaning … Human cognition and understanding are bypassed as irrelevant to objective meaning relations. All mental processes (ideas, images, imaginative projections) that might explain how it is that a sign could come to connect up with the world, and with other signs, are excluded from consideration. This is the Objectivist view of meaning in its purest form. (Johnson 1987: xxx–xxxi)

In response, Itkonen points out that Frege by no means "bypasses" perception, action and even emotion, as obvious from what he write on "ideas", but that his "anti-psychologism", which is very similar to that of Husserl (2001 [1900]) concerns "the common property of many": the level of intersubjective, common knowledge, which is normative, and accessible through intuition (section on Metaphysics). It is hardly correct to say that this "level" does not involve "human cognition and understanding", since that would make it indeed Platonist.[7] Rather, the challenge is to articulate how the "shared, common mind" is constructed through processes and structures of human consciousness, which are different from those involved in individual

---

[7]This is a mistake committed by Katz (1981): "The properties Katz assigns to abstract objects appear all to be possessed by the kind of conventions of mutual knowledge that Esa Itkonen argues are constitutive of linguistic rules (Itkonen 1978; not cited in Katz 1981)" (Pateman 1987: 2).

experience. Common knowledge, Itkonen proposes, can be viewed as a *network*, which is not reducible to the individual pieces of thread, but is nevertheless composed by them once they come to stand in a particular structure (Itkonen 2008). Therefore, Itkonen defends a "social ontology" of meaning, which, as stated in section on Metaphysics is consistent with (the later) Husserl's emphasis on intersubjectivity and the personalistic attitude, revealing "the intersubjective constitution of the world".

More recently, this "minority position" within CL, represented by Itkonen, Sinha, Zlatev and a few others, has gained some prominence, and "intersubjective" is becoming a commonplace term. Verhagen (2005, 2008) has argued that the fundamentally intersubjective nature of language is reflected in the fact that argumentative and evaluative meanings are lexicalized and codified in grammar, and thus part of conventional meaning (semantics) and not only context-specific meaning (pragmatics). Several authors in the edited volume *The Shared Mind: Perspectives on Intersubjectivity* (Zlatev et al. 2008), furthermore explore how more basic (pre-linguistic and pre-normative) processes of intersubjectivity are involved in the *emergence* of shared meaning (as well as the ego itself) from developmental and evolutionary perspectives (Gallagher and Hutto 2008; Hutto 2008; Zlatev 2008), thereby showing how conceptual/phenomenological and empirical investigations could be interrelated. The "hierarchical layerings (*Stufenbau*)" of Husserl's *genetic phenomenology* (cf. Moran 2005: 218) seem to have a close affinity with the "layered model" of the *sense of self* of Daniel Stern (2000 [1985]), which has been influential for several CL-related authors' work, such as the *mimesis hierarchy* proposed by Zlatev (2008). Stern's method of combing empathetic understanding based on primary interaction with children, with the more objective methods of developmental psychology, revealing e.g. proto-conversations (Trevarthen 1979) and neonatal imitation (Meltzoff and Moore 1983) is also a good example of how the personalistic and psychological attitudes may complement each other, rather than stand in conflict. Gallagher (2005) has contributed with phenomenological concepts such as "agency", "ownership", "body schema" and "body image" to these investigations, linking intersubjectivity to a theme that is central for both phenomenology and CL: embodiment.

Embodiment

The concept of *embodiment* has become a central notion not only in CL (Lakoff 1987; Johnson 1987; Zlatev 1997; Evans 2003), but in cognitive science (Varela et al. 1991; Clark 1997) and the neuroscience of consciousness (e.g. Edelman 1992; Damasio 2000). It is even claimed to unite efforts in these and other related fields into what is sometimes called "second generation cognitive science" (Gallese and Lakoff 2005). However, due to the ambiguity or the term, demonstrated by a recent interdisciplinary collection on the topic (Ziemke et al. 2007), this project of unification is unlikely to succeed.

Zlatev (2007) argues that most "embodiment theorists", including those within CL, undervalue concepts which are central for phenomenology such as *normativity*

(see section on Methodology: Phenomenological and "Empirical" Methods), *representation*, and *intentionality*. Since these are essential concepts for any theory of language, it is unclear whether an "embodied cognitive science" has the conceptual tools to provide any coherent explanation of language. This controversial claim is substantiated by examining the most authoritative presentation of this program, from the standpoint of CL (Lakoff and Johnson 1999).

Lakoff and Johnson propose that "…there are at least three levels to what we are calling the embodiment of concepts: the neural level, phenomenological conscious experience and the cognitive unconscious" (ibid: 102). Starting from the bottom, "neural embodiment concerns structures that characterize concepts and cognitive operations at the neural level" (ibid: 102). It is claimed that this level "significantly determines […] what concepts can be and what language can be" (ibid: 104). One of the most specific definitions of "an embodied concept" is provided in terms of this level only: "*An embodied concept is a neural structure that is part of, or makes use of the sensorimotor system of our brains. Much of conceptual inference is, therefore, sensorimotor inference*" (ibid: 20, *original emphasis*). The "neural level" refers furthermore to a higher-level description that is heavily dependent on "an important metaphor to conceptualize neural structure in electronic terms" (ibid: 103). The connectionist model of Regier (1996) is given as an instance of "neural modeling", even though it is only inspired in the most general terms from what is known about the brain.

The next level, referred to as *phenomenological embodiment*, is devoted much less attention. Its first definition is "[…] the way we schematize our own bodies and things we interact with daily" (ibid: 36), with reference to the phenomenological tradition and specifically the work of Gallagher (cf. Gallagher 2005). The second definition is considerably broader, and seems to overlap with Husserl's Lifeworld: "It (i.e. phenomenological embodiment) consists of *everything we can be aware of*, especially our own mental states, our bodies, our environment and our physical and social interactions. This is the level at which we speak of the "feel" of experience […]" (ibid: 103, emphasis added). At the same time, Lakoff and Johnson claim that "phenomenology also hypothesizes nonconscious structures that underlie and make possible the structure of our conscious experience" (ibid: 103). This introduces the main operative level in Lakoff and Johnson's account of embodiment:

> The cognitive unconscious is the massive portion of the iceberg that lies below the surface, below the visible tip that is consciousness. It consists of all those mental operations that structure and make possible all conscious experience, including the understanding and use of language. (ibid: 103)

This level is said to be "the realm of thought that is completely and irrevocably inaccessible to direct conscious introspection" (ibid: 12) and (nearly) all-pervasive: the cognitive unconscious constitutes "the 95 percent below the surface of conscious awareness [that] shapes and structures all conscious thought" (ibid: 13). Lakoff and Johnson claim that it is "hypothesized on the basis of convergent evidence, […] required for scientific explanation" (ibid: 115) and that "the detailed processes and structures of the cognitive unconscious (e.g., basic-level categories, prototypes, image schemas, nouns, verbs, and vowels) are hypothesized to make sense of conscious

behavior" (ibid: 104). So it turns out that this all-important level of embodiment is a phenomenologically inaccessible theoretical construct.

Again, this is hardly consistent with phenomenology. First, "nonconscious" forms of intentionality are indeed recognized by Husserl, as well as by his successors, but these are not "completely and irrevocably inaccessible", as stated by Depraz (2001: 171–172):

> Such a functional (*fungierend*) habitual intentionality includes a part of automaticity, what Husserl calls individual *Habitualität* and collective *Sedimentierung*; but unlike what happens at the neurocomputational level, where neural processes are to be found, I have the ability to *reactivate*, as Husserl says, such an unconscious functionality so as to become aware of it. (ibid: 171–172)

Second, as pointed out earlier, phenomenological analysis is not equivalent to "direct conscious introspection". Finally, many of the phenomena placed in the "cognitive unconscious" (e.g. "nouns, verbs and vowels" are clearly accessible: through linguistic analysis based on native speaker intuitions in the manner shown in section on Methodology: Phenomenological and "Empirical" Methods.

It is clear that Lakoff and Johnson (1999) feel pressed to defend the "reality" of the "cognitive unconscious" and they attempt to do so repeatedly. Perhaps the most revealing statement is "To say that the cognitive unconscious is real is very much like saying that neural "computation" is real" (ibid: 104). But is neural computation "real"? Within information-processing, "classical" cognitive science a common way to make the distinction between mental and non-mental without recourse to consciousness is through the notion of *computation*: mental processes are involved in (symbolic) computation, non-mental ones are not (e.g. Pinker 1994). Despite their overall opposition to information processing theorists, through their endorsement of "neural computation" Lakoff and Johnson come surprisingly close to the position of "first generation" computational cognitive scientist. Furthermore the "computational" solution to the mental/non-mental distinction does not work for a very simple reason: there is no *intrinsic* computation going on in the brain, as argued at length by e.g. Searle (2002). All talk of neural computation is metaphorical, in the sense that it is a matter of attribution from the outside. And because of that, the "computational level" is not ontologically or causally distinct from the neural level: "Except in cases where an agent is actually intentionally carrying out a computation, the computational description does not identify a separate causal level distinct from the physical structure of the organism" (Searle 2002: 126).[8]

---

[8]A defense of the reality of "unconscious computation", suggested by an anonymous reviewer, in line with the view of metaphor expressed by Lakoff and Johnson (1999) is that what is termed "metaphorical" is no less real than what is termed "literal", as long as it plays an explanatory role. And the level of the "cognitive unconscious" plays such a role, both upwards towards consciousness and downwards with respect to the brain. However, this seems to conflate ontology and epistemology (cf. footnote 3). While both consciousness and the brain are in the broad phenomenological sense *empirical* phenomena, "unconscious computation" is not, and simply a form of *describing* either one or the other. But in attempting to capture aspects from both – intentionality on the one hand, and bio-physical causality on the other – it becomes simply incoherent, which is why Searle (1992) argues that we would do best to dispend with it. For a more extensive argument, cf. Zlatev (2007a).

In sum, Lakoff and Johnson's (1999) concept of embodiment is obviously inconsistent with a phenomenological view of language, or indeed with any view emphasizing the properties of normativity (conventionality), representation and intentionality (Zlatev 2007). Regarding the first, there are frequent references to "conventional mental imagery" (ibid: 45), but it is not even made clear whether this imagery is conscious or only part of the "cognitive unconscious" – not to mention the question of how this imagery would be shared, and furthermore known to be shared, which is necessary for it to have a normative component (see section on Methodology: Phenomenological and "Empirical" Methods). One could say the same for the use of the term "conventional metaphor" in the cognitive linguistic literature – there is nothing "conventional" about neurally realized domain-to-domain mappings, at least in any conventional use of the term 'convention' (e.g. Itkonen 2003; Zinken 2007). The most Lakoff and Johnson can do to account for shared meanings is, as mentioned earlier, to point out that "commonalities […] exist in the way our minds are embodied" (ibid: 4). Concerning the concept of representation, Lakoff and Johnson are explicitly antithetical.

> As we said in *Philosophy in the Flesh*, the only workable theory of representations is one in which a representation is a flexible pattern of organism-environment interactions, and not some inner mental entity that somehow gets hooked up with parts of the external world by a strange relation called 'reference'. We reject such classical notions of representation, along with the views of meaning and reference that are built on then. Representation is a term that we try carefully to avoid. (Johnson and Lakoff 2002: 249–250)

A similar, if not stronger, form of anti-representationalism (and anti-intentionality, see below) is advanced by Johnson and Rohrer (2007: 30): "According to our interactionist view, [neural] maps and other structures of organism-environment co-ordination are prime examples of non-representational structures of meaning, understanding, and thought." In their urge to dissociate themselves from any "dualism", scholars like Lakoff, Johnson and Rohrer, adopt a reductive "monism", where consciousness, representation, reference, intentionality all give way to all-pervading "organism-environment interactions". It is one thing to (justly) argue against "representations" or "signs" in perception and active involvement, as common within the phenomenological tradition, and quite another to deny that, say, a picture is a representation of whatever it depicts, irrespective of whether the latter exists in the "real world" or not (Sonesson 2007). It is in this latter sense that some, though not all, language use is representational (Zlatev 2007). Furthermore, to deny that assertions are a kind of representation is to deny for example that a description of a situation can be either true or false. Lakoff and Johnson (1999) would be inconsistent to deny this since in their own definition of "embodied truth" a person holding a sentence to be "true" is said to understand the sentence to "accord" with "what he or she understands the situation to be" (ibid: 106), i.e. "a kind of cognitivist correspondence theory of truth (Haser 2005: 84), or internal realism (see section on Metaphysics).[9]

---

[9]In the name of fairness, it should be pointed out that Lakoff, Johnson and Rohrer do not explicitly deny a representational relation between language or pictures and reality, but of "internal representations" of the kind assumed by "first generation" cognitive scientists. However, they never

Finally, and perhaps most damaging for any meaningful relation between (most) "embodiment theories" in CL and phenomenology – despite cursory and misleading references to "phenomenological embodiment" – is that not only is there no room left for representation or signs (*picture consciousness*, in the sense of Husserl, see section on Representation and Sign), but for *intentionality*. It is true that Johnson (1987) repeatedly emphasized its importance, with references to Searle (1983) if not Husserl, and in *Philosophy in the Flesh*, some of the clearly non-neural structures of the "cognitive unconscious" (e.g. the concept CHAIR) are said to be operating with some sort of "unconscious intentionality". But if this is "completely and irrevocably inaccessible" to consciousness, then this notion does not make sense from a phenomenological point of view, as argued earlier. Finally, in their most recent writings, the main "experientialist" philosophers only speak of undirected "organism-environment interactions", since any talk of intentionality would imply a form of "dualism", and this is rejected with hardly any argument in the name of "embodiment".

For lack of space, I do not delve into the radically different concept of embodiment in phenomenology (e.g. Thompson 2001; Sonesson 2007; Gallagher 2005) or the work of Zlatev (2005, 2007, 2008) on the role of *bodily mimesis*, conscious bodily representational movement and imagery, for bridging the gap between individual bodily experience and language. I will only mention that such a conception ties naturally the notions of embodiment and intersubjectivity, which as pointed out by Zahavi (2001) was one of Husserl's crucial insights:

> My body is given to me as an interiority, as a volitional structure, and as a dimension of sensing, but it is also given as a visually and tactually appearing exteriority. But what is the relation between that which Husserl calls "Innen-" and the "Aussenleiblichkeit", i.e. what is the relation between the lived bodily inwardness on the one hand, and the externality of the body on the other? (Husserl 1973c, p. 337) /.../ I am experiencing myself in a manner that anticipates both the way in which an other would experience me and the way in which I would experience an other. This may be what Husserl is referring to when he writes that the possibility of sociality presupposes a certain intersubjectivity of the body. (Zahavi 2001: 161).

Summary

This section has reviewed discussions within CL with respect to foundational issues which are also central for phenomenology. We may conclude that despite some superficial similarities, there is a sizable gulf between central conceptions concerning ontology, method, language, consciousness, body and intersubjectivity held by the "mainsteam" representatives of CL, and those of phenomenology. Furthermore, it seems that this gulf has been widening over the past two decades, at least in the

---

provide an account of "external representations" either, or of internalized such (Vygotsky 1978), and fail to draw a distinction between sensory-motor and representational processes (cf. Ikegami and Zlatev 2007), due to their insistence on a strong form of "evolutionary continuity".

writings of Lakoff and Johnson. From a Husserlian perspective, such a perspective suffers from a bi-polar abstraction from the multi-perspectival Lifeworld: the psychological attitude and the biological attitude. To avoid contradiction and "dualism" these are furthermore collapsed, resulting in an ontology without room not only for normativity (which was absent from the start), but even for intentionality and representations/signs. Methodologically, intuition and introspection are not distinguished, and considered at best as "evidence" on a par with that of the natural sciences. I believe that this is an attitude that may even be called "anti-phenomenological".

On the other hand, the views of critics such as Itkonen and myself, representatives of a "minor stream" within CL, were shown to be consistent with phenomenology. There are also indications that this position is increasing in influence (though of course, it is difficult for me to be objective in this case), as shown in some recent publications (Popova 2005; Croft 2007; Zlatev et al. 2008). There is even a tendency among some of the mainstream authors to "get back to experience". In a recent publication, Johnson (2005: 28) writes: "Meaning is a matter concerning how we understand situations, people, things and events, and this is as much a matter of values, felt qualities, and motivations, as it is about structures of experience". In any case, it is clear that a rapprochement between phenomenology and CL is under way, and is bound to play a major role in future debates on fundamental issues such as those discussed in this section. In the next section, I focus instead on more specific issues, where such a rapprochement already has had a beneficial influence on CL.

Phenomenological Influences

Representation and Sign

The concept of *representation* is one of the fundamental, and at the same time one of the most ambiguous ones in cognitive science. The same can be said about the concept *sign* in the various schools of semiotics (Cassirer 1957; Eco 1982; Lotman et al. 1975). Sonesson (2006, 2007) furthermore remarks on the curious parallel current tendency in both traditions to reject these fundamental concepts. But this is premature, since they are seldom given a clear explication, and if so defined differently by different authors. Thus it is unsurprising that a rational debate on their nature and role in (human) cognition and consciousness has been difficult. Still, without them a phenomenologically (and logically) valid account of e.g. language would be impossible (see section on Embodiment).

Husserl, and phenomenology in general, as well as Gestalt psychology (Gurwitsch 1964) and Gibsonian ecological psychology (cf. Costall 2007) repudiate an analysis of *perception* in terms of "mental representations", since among other things, this leads to a form of subjectivism about the mind (combined with objectivism about the world). But this does not imply that there are no representations in consciousness in general. "One of Husserl's most significant achievements

is to show that picture-consciousness (*Bildbewusstsein*) is a specific modality of consciousness. The error of modern philosophy had been to misconstrue perception and imagination as picture-consciousness." (Moran 2005: 169). In seeing a painting of, e.g. an apple, we see both the painting directly (*Bildding*), and the content represented (*Bildobjekt*) indirectly. It is the content that is the *theme,* not the actual stokes of paint and canvas of the *Bildding*. Both are to be distinguished from the actual apple (if such exists at all) that has been painted (*Sujet*).[10] Sonesson (1989, 2007) uses this analysis, along with Husserl's notion of *appresentation*, as the basis for his influential definition of the sign-concept in what has been called "phenomenological semiotics". He combines Husserl's insight with a Piagetian conception of the "sign function": "there is a *differentiation* between expression and content in the double sense … that they *do not go into each other in time and/or space*, and they are perceived to be *of different nature* (Sonesson 2007: 93, original emphasis). Taken together this implies that if a given subject (say, an infant or a non-human primate) would attempt to eat the painting of the apple, he would not be seeing it as a *sign*, since he would be collapsing the distinctions between *Bildding, Bildobjekt* and *Sujet*. Ikegami and Zlatev (2007: 200) propose defining *representation* similarly:

> A representation is a structure that consists of three parts: an *expression* that stands for a given *content* for a given *subject*. Thus defined, it is identical with the classical definition of a *sign. …* A clear example of a representation is a picture: the depicted apple cannot be eaten, but it represents (in this case *iconically*) an apple that can.

Imagination and language can thus be said to involve representations, but (as pointed out by Husserl) also a different form of consciousness than picture-consciousness. In the case of imagination (fantasy), there is no directly perceived, unthematic *Bildding*, though still there is a differentiation between the mental image and the real situation. Language can be said to involve "*Bildding"* (though Husserl's would not have used this analogy): the spoken, written or signed expression. But what is most often thematic is not the content (sense) but the referent ("*Sujet*"). Furthermore, while in many cases there may be an associated "image" associated with the expression (cf. the quote from Frege in section on Intersubjectivity), the sense of the expression cannot be identified with this, but rather with a *shared convention for the correct application* of the expression (cf. section on Intersubjectivity). Finally, while pictures represent iconically (based on a similarity-relation between expression and content), languages does so "symbolically", i.e. based on a *conventional* (though not necessarily "arbitrary") relation between expression and content.

Such phenomenological-semiotic analyses are prerequisites for adequate empirical investigations, and indeed they have been used in developmental and evolutionary studies of transitions "from pre-representational cognition to language". For example,

---

[10] Sonesson (1989) has, however, shown that this picture needs to be complicated a bit. *Bildobjekt* cannot be the mediating content, since the latter involves a sort of "quasi-perception" in the words of Husserl, e.g. in perceiving a black and white photo, the skin of a given person is seen as white, or grey, while I know that the actual *Sujet* (referent) has a certain color. Hence, Sonesson adds another mediating "layer", corresponding to, in the example given, to *the apple, as I know that it is: colorful, round etc*, that would continue to exist even if the last apple on Earth were eaten.

it is the major hypothesis of the project *Stages in the Evolution and Development of Sign Use* (Zlatev and SEDSU-Project 2006), that what makes human consciousness different from that of other primates is above all due to one or more aspects of the *sign-function* (in the sense defined in the previous section), rather than to language per se. Indeed in the experiments carried out it has been shown that even great apes have enormous difficulties in dealing adequately with pictures. Piaget (1945) furthermore, suggested that the sign-function emerges in children at the end of the sensory-motor period around 18 months, by generalizing and internalizing acts of *imitation*. The latter involves the ability to both match and differentiate between one's own body (image) and the model, and is thus a likely precursor to both picture-consciousness and language. Moreover, it may serve as the basis for the emergence of *socially-shared* meanings, since imitation implies a basic form of intersubjectivity (Zlatev 2005, 2007, 2008). Empirical studies investigating this hypothesis are under way.

Image Schemas

The concept of *image schema,* introduced by Johnson (1987; see Section on Methodology: Phenomenological and "Empirical" Methods), is central to CL. At the same time, as shown by the chapters in the volume edited by Hampe (2005), there is little agreement on its meaning. Most often image schemas are thought to be rather abstract structures such as PATH and VERTICALITY (Johnson 1987; Mandler 2004), or even more abstract ones such as CYCLE and PROCESS (cf. Grady 2005). On the other hand, sometimes "basic level" experience-types such as PUSH and GRASP are given as illustrations of image schematic structure (Gibbs 2005). While some authors characterize these as representational structures (Lakoff 1987; Grady 2005; Mandler 2004), others emphasize their non-representational, "interactional" character (Johnson and Rohrer 2007; Johnson 2005; Gibbs 2005). Some see them as part of the "cognitive unconscious" (Lakoff and Johnson 1999; Johnson 2005; cf. Section on Embodiment), but others claim that they possess phenomenal contours and hence cannot be completely unconscious (Gibbs 2005). Finally, even their "embodied" nature, in the sense of being based on physical experience, has been questioned in some definitions (Clausner and Croft 1999), while Grady (2005) argues for limiting the notion strictly to sensorimotor experience. Recently, Langacker (2006: 36) has proposed a novel, and rather interesting, interpretation of the concept as "subjectively construed mental operations" (see below).

Clearly, therefore, these are not one but a number of *different* concepts. Zlatev (2005) analyzes these differences along the "parameters": representational status, accessibility to consciousness, level of abstractness, sensory modalities, and (inter)subjectivity. In the section on Fundamental Issues: "Experientialism" vs. Phenomenology, some of the non-phenomenological interpretations of the concept where subjected to criticism, which I will not repeat. Instead, I focus on a few

interpretations which appear to be more or less congruent with phenomenology, i.e. those which view image schemas as *structures or processes of consciousness*.

Grady (2005) proposes that there are at least three different levels of abstractness involved: (i) concrete schemas such as UP, for which Grady reserves the designation "image schema", (ii) more abstract schemas such as MORE, which Grady calls "response schemas" since he regards them as being the outcome of *primary metaphors*, mappings from concrete domains due to correlations in experience in early childhood, and (iii) "superschemas" such as SCALAR PROPERTY, which capture the shared structure between (i) and (ii), guaranteeing a degree of isomorphism. This hierarchy is useful, but Grady fails to specify the nature and origin of such superschemas providing the constraints on "metaphorical" mappings. Dewell (2005: 388) argues that schemas such as CONTAINMENT and UP are not "purely preverbal" but rather shaped by language itself: "the influence of language generally toward maximally precise and differentiated linear shapes that can be explicitly profiled and publicly accessed from a flexible perspective." But if even the most concrete schemas that Grady presents are language-based at least to some degree, there is perhaps even more reason to believe that this is the case with the more abstract ones. Superschemas such as ONTOLOGICAL CATEGORY (EVENT, PROCESS, THING), SCALARITY/DIMENSIONALITY, ASPECT, BOUNDEDNESS etc. are all reflected in the grammatical systems of the worlds' languages. Does such universality imply the need for an explanation in terms of "preverbal" structures and processes as is customary in CL?

Not necessarily. As shown by Heine and Kuteva (2002) processes of grammaticalization can lead not only to language change, but to an increase in language complexity through the *evolution* of grammatical (and hence semantic) categories. As for the "metaphorical mappings" between more concrete image schemas and what Grady calls "response schemas", these could possibly be explained as deriving from (conscious) processes of *analogy* (Itkonen 2005), performed by speakers under the constraints of the shared structure (the "superschemas") in the source and target domains. Such an approach would imply a rather strong role of language on human consciousness.

Another, more consciousness-first approach, at least for the lower level of Grady's hierarchy, is suggested by the developmental psychologist Mandler (2004) who presents a theory in which infants are neither sophisticated pattern learners (as in most connectionist models) nor driven by unconscious innate "domain-specific" knowledge and language-acquisition devices, but conscious beings who construct their conceptual systems piece by piece, above all through the key process of *perceptual meaning analysis*, through which infants "consciously analyze what objects are doing. The result of this process – interpretations of the world that suffuse the mind with meaning – are also accessible to consciousness" (ibid: 292). Mandler sees image schemas like PATH and CONTAINMENT, which she defines as "analog representations that summarize spatial relations and movements in space" (ibid: 79) as resulting from this process. But in what sense are they "accessible to consciousness"? Mandler tries to answer this troublesome question by calling image schemas "conceptual/representational format" which as such is not conscious,

while its *content* is consciously accessible. However, what this distinction amounts to is not clear.

On the basis of extensive experience with semantic analysis, employing the concepts of *construal* and *profiling*, as part of his theory of Cognitive Grammar, Langacker (2006: 36) proposes a rather original shift in perspective:

> [W]e should not think of image schemas as something we conceptualize (which the term *image* might suggest), but as *cognitive abilities inherent in the conception of other entities*. For instance, the *source-path-goal* image schema could instead be thought of as the capacity for mental scanning. The *link* schema could be thought of as the capacity to exploit a conceptual connection. The *centre-periphery* schema might be thought of as an asymmetry in mental access… Mandler's (1991) notion of perceptual [meaning] analysis – the redescription of sensorimotor experience in image schematic form – could then be explicated as the apprehension of primary experience by means of such processing capacities (presumably inborn). (emphasis added)

In other (and phenomenological) words, Langacker is proposing to regard "image schemas" not as structures, representations etc., but as *processes* of (human) consciousness through which experience is "analyzed", and which therefore, in line with the nature of *the natural attitude* (cf. section on Metaphysics), are in the background of consciousness, or as Langacker says "subjectively construed". However, since consciousness is always pre-thematically aware of itself, "this pre-reflective awareness can be turned into explicit reflection. Indeed, it is one of Husserl's eidetic laws that every mental process is so structured that one can turn one's gaze on it and identify its components (Ideen I #98, p.241)" (Moran 2005: 145). Thus, these processes can be investigated through eidetic variation, and not simply through "informal phenomenological analysis" (Johnson 2005). However, whether they are "inborn" (as Langacker and Mandler suppose), acquired preverbally, as Johnson hypothesises, or in some cases even acquired "post-verbally" as suggested above, is not something that phenomenology can answer, but rather an open question for empirical psychology. This is a good example of the complementary relation between the two.

Construal

In CL in general, and in Cognitive Grammar (Langacker 1987) in particular, the notion of construal plays a central role. In its more general sense, it refers to "a range of cognitive processes" (Saeed 2003: 376) mediating between "objective reality" and linguistically expressed meaning, thus highlighting CL's position that "meaning resides in conceptualization". On the face of it, this seems to entail a subjectivist theory of meaning, facing phenomenological, Wittgensteinian and Fregean objections that linguistic meanings are necessarily intersubjective (see section on Intersubjectivity). However, there may be a way to resolve this contradiction, given a phenomenological take on the notion of construal. In a recent publication (the same in which "image schemas" are analyzed as mental processes referred to
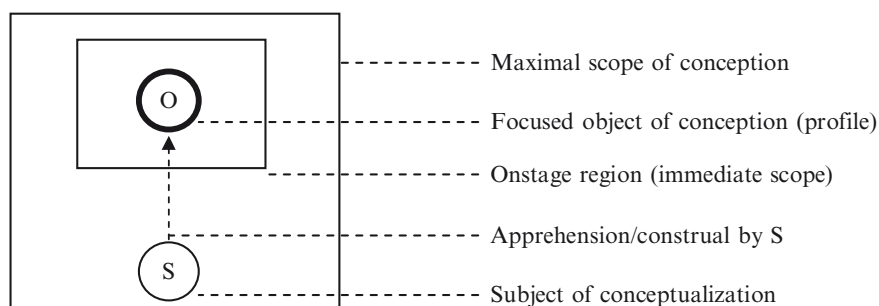
Fig. 1 Subjective and objective construal (adapted from Langacker 2006: 19)

above) (Langacker 2006: 18) explains the difference between "objective" and "subjective" construal as follows:

> An entity is said to be *objectively construed*, to the extent that it goes "onstage" as an explicit, focused object of conception … An entity is *subjectively construed*, to the extent that it remains "offstage" as an implicit unselfconscious subject of conception. At issue, then, is the inherent asymmetry between the conceptualizer and the conceptualized, between *the tacit conceptualizing presence and the target of conceptualization*. (emphasis added)

It is easy to see this as a paraphrase of a phenomenological analysis of consciousness: the "objectively construed" is the *theme*, while the whole "onstage" region is the *field of consciousness* (Gurwitsch 1964). The "conceptualizer", situated in the *background* (*margins*) of consciousness ("offstage") is the ego, and the asymmetric relationship between the latter and the "conceptualized" is that of *intentionality* itself. In line with Gurwitsch's claims that "this thematic structure translates to language" (Sonesson 2007: 107), Langacker implies that all language use, and conventional linguistic meanings, utilize this structure (see Fig. 1)[11]:

> In principle, an expression's conceptualizing meaning always incorporates the conceptualizing presence who apprehends and construes the situation described. … Minimally, subjectively construed elements *include the speaker, and secondarily the addressee, who employ the expression and thereby apprehend its meaning*. Minimally, objectively construed elements include the expression's *profile*, i.e. what it designates (or refers to) within the conception evoked. (Langacker 2006: 18, emphasis added)

The novel aspect in comparison to earlier formulations is that here Langacker includes *both the speaker and the addressee* (the latter not included in the diagram) and, sympathetically interpreted, their joint "apprehension of the meaning" of the expression in the background of consciousness, i.e. the (normative) expectation of its shared meaning. This interpretation highlights both the continuity between language and perceptual consciousness (both involving intentionality and thematic structure), and some of its difference: the second is necessarily intersubjective

---

[11] The unacknowledged parallels between Gurwitsch and Langacker are further explored by Sonesson (2004).

(though at least in Husserl's later thought, the "world" itself is "constituted" through intersubjectivity, see section on Metaphysics). What is still lacking for language, though, is a place of the *sign* (see section on Representation and Sign) in this analysis: Langacker writes of structures such as those in Fig. 1 as the "semantic pole", while expressions are the "phonological pole" of signs. But clearly, the latter cannot be seen as external to thematic structure. It could perhaps be included in the "offstage" region in Fig. 1, between conceptualizer and the "onstage" (thematic) region.

Langacker proceeds to explain how this analysis can be applied to the phenomenon of *grammaticalization*, i.e. historical change whereby the meaning of lexical items becomes "bleached" in order to turn into a grammatical morpheme, e.g. the verb *going to* from verb of motion to future tense marker, as a process of "subjectification": a shift from the "onstage" to the "offstage" region. This is a more subtle analysis than simply stating that the process is one in which meanings become more "subjective" (Traugott 1989). The details of this are too complex to address here, but let us conclude by suggesting that if the phenomenological interpretation of Langacker's notion of construal suggested here is adequate, subjectification, and Cognitive Grammar analyses of a large number of semantic phenomena such as quantification, which are quite different from analyses within e.g. Montague semantics (Montague 1974), could be regarded as the fruits of a "phenomenological linguistics" (cf. Zlatev 2008).

Conclusion

In this chapter, I have argued that while the explicit philosophy and the meta-theory of "mainstream" CL are largely incongruent with phenomenology, the thinking of less "prototypical" representatives of CL such as Itkonen, Sinha, Harder and Zlatev, and perhaps more importantly: the *practice* of CL has some considerable overlaps with phenomenology. These deserve to be further explored. The incongruency between "experientialism" and phenomenology seems to be largely due to a degree of philosophical naiveté (or a particular rhetorical style) within the first (cf. Haser 2005), combined with a mentalist/subjectivist bias, inherited from the Chomskyan school, from which CL emerged at the same time as it rebelled against it (cf. Itkonen 2003). Instead of attempting to overcome this through a sociocultural (Wittgensteinian, Vygotskyan or phenomenological) perspective, the opposite tendency of a more highly pronounced bio-physical attitude is observed in the writings of some prominent representatives (e.g. Dodge and Lakoff 2005). Consciousness, intentionality, and even representation/sign are downplayed, if not outright rejected within this attitude. When "phenomenology" is acknowledged ("Cognitive semantics is thus a branch of phenomenology" (Talmy 2000: 4)), this is interpreted subjectively, thus bringing the circle back to the beginning – and provoking new reactions from those who wish to "naturalize" language and linguistics by applying *only* "empirical methods" (cf. Introduction).

I would venture to propose that CL is currently at a crossroads: if this 30-year old tradition is to break from the vicious circle of subjectivism-physicalism, it would need to seriously reconsider its foundations, and account for concepts such as intersubjectivity, normativiy, sign-function, and consciousness, which are essential for language (Zlatev 2007). A natural way for it to do so is by making a rapprochement with phenomenology. In the second part of this review, I have focused on some "signs" that such a rapprochement is not only possible, but under way. Space does not allow me to review some more specific analyses, in which phenomenological concepts or themes have been fruitfully applied: the analysis of "the phenomenology of negation and its expression in language" (Saury 2004), "the origins of grammar in the verbalization of experience" (Croft 2007), the role of the sense of touch in "synaesthetic metaphors" (Popova 2005) and the verbalization of "motion situations" (Zlatev et al. in press).

Finally, what Husserl called "correlation research" into different aspects of the Lifeworld, as opposed to dogmatic metaphysical "absolutizations" seems to be urgently required in order to counter the dominant present day physicalist (and calculating) attitude, as well as to mend the rift between different kinds of research in language. On the one hand, traditional, intuition-based linguistics is based on "hermeneutic" methods such as explication (Itkonen 1978). On the other hand, psycho- and neurolinguistics study causal processes involved in, e.g. language acquisition, production and comprehension, using mechanistic (consciousness-independent) models. But if consciousness is essential for the *existence* of language, its analysis, as well as for its acquisition and use (Zlatev 2008), then a phenomenological perspective on language may help unite (in a non-reductive way) traditional and "empirical" linguistics in a truly *general* (cognitive) linguistics. Non-reductive projects of "naturalization" within phenomenology (e.g. Gallagher and Broested-Soerensen 2006) can serve as an example of how this could be carried out in practice.

References

Cassirer E (1957) The philosophy of symbolic forms. III. The phenomenology of knowledge. Yale University Press, New Haven

Chomsky N (1975) Reflections on language. Pantheon, New York

Clark A (1997) Being there: putting brain, body, and world together again. MIT Press, Cambridge, MA

Clausner T, Croft W (1999) Domains and image schemas. Cogn Linguist 10:1–32

Cohen JL (1986) The dialogue of reason: an analysis of analytical philosophy. Clarendon, Oxford

Costall A (2007) Bringing the body back to life: James Gibson's ecology of embodied agency. In: Ziemke T, Zlatev J, Frank R. (eds) Body, language and mind. Embodiment Mouton de Gruyter, Berlin 1:55–83

Croft W (2007) The origins of grammar in the verbalization of experience. Cogn Linguist 18(3):339–382

Damasio A (2000) The feeling of what happens. body, emotion and the making of consciousness. Harvester, New York

Dennett D (1987) The intentional stance. MIT Press, Cambridge, MA

Depraz N (2001) The Husserlian theory of intersubjectivity as alterology. J Conscious Stud 8(5–7):169–178

Dewell R (2005) Dynamic patterns of CONTAINMENT. In: Hampe, B (ed) From perception to meaning. Image schemas in cognitive linguistics. Mouton de Gruyter, Berlin, pp 369–394

Dewey J (1981 [1925]) John Dewey, The later works, 1925–1953, vol 1. Southern Illinois University Press, Carbondale

Dodge E, Lakoff G (2005) Image schemas: from linguistic analysis to neural grounding. In: Hampe B (ed) From perception to meaning: image schemas in cognitive linguistics. Mouton de Gruyter, Berlin, pp 57–91

Eco U (1982) Semiotics and the philosophy of language. Indiana University Press, Bloomington, IN

Edelman G (1992) Bright air, brilliant fire: on the matter of the mind. Basic Books, New York

Evans V (2003) The structure of time. language, meaning and temporal cognition. Benjamins, Amsterdam

Fauconnier G, Turner M (2002) The way we think: conceptual blending and the mind's hidden complexities. Basic Books, New York

Frege G (1966 [1892]) On sense and referece. In: Geach P, Black M (eds) Translations from the philosophical writings of Gottlob Frege. Basil Blackwell, Oxford

Gallagher S (2005) How the body shapes the mind. Oxford University Press, Oxford

Gallagher S, Broested-Soerensen J (2006) Experimenting with phenomenology. Conscious Cogn 15:119–134

Gallagher S, Hutto DD (2008) Understadning others through primary interactions and narrative practice. In: Zlatev J, Racine T, Sinha C, Itkonen E (eds) The shared mind: perspectives on intersubjectivity. Benjamins, Amsterdam, pp 17–38

Gallese V, Lakoff G (2005) The brain's concepts: the role of the sensory-motor system in conceptual knowledge. Cogn Neuropsychol 22(3/4):455–479

Gardner H (1987) The mind's new science. Basic Books, New York

Geeraert D, Cuyckens H (eds) (2007a) The Oxford handbook of cognitive linguistics. Oxford University Press, Oxford

Geeraerts D, Cuyckens H (2007b) Introducing cognitive linguistics. In: Geeraerts D, Cuyckens H (eds) The Oxford handbook of cognitive linguistics. Oxford University Press, Oxford, pp 3–21

Gibbs R (2005) The psychological status of image schemas. In: Hampe B (ed) From perception to meaning: image schemas in cognitive linguistics. Mouton de Gruyter, Berlin, pp 113–135

Gibson J (1979) The ecological approach to visual perception. Houghton Mifflin, Boston, MA

Grady JE (2005) Image schemas and perception: refining the definition. In: Hampe B (ed) From perception to meaning: image schemas in cognitive linguistics. Mouton de Gruyter, Berlin, pp 35–56

Gurwitsch A (1964) The field of consciousness. Duquesne University Press, Pittsburgh

Hampe B (ed) (2005) From perception to meaning: image schemas in cognitive linguistics. Mouton de Gruyter, Berlin

Harder P (2007) Cognitive linguistics and philosophy. In: Geeraerts D, Cuyckens H (eds) The Oxford handbook in cognitive linguistics. Oxford University Press, Oxford, pp 1241–1265

Haser V (2005) Metaphor, metonymy and experientialist philosophy. Mouton de Gruyter, Berlin

Heine B, Kuteva T (2002) On the evolution of grammatical forms. In: Wray A (ed) The transition to language. Oxford University Press, Oxford, pp 376–397

Honderich T (2006) Radical externalism. Journal of Consciousness Studies 13(7–8):3–13

Husserl E (1970 [1936]) The crisis of European sciences and transcentental phenomenology. Northwestern University Press, Evanston, IL

Husserl E (1977 [1925]) Phenomenological psychology. Nijhoff, The Hague

Husserl E (1989 [1952]) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy, second book. Klewer, Dordrecht

Husserl E (2001 [1900]) Logical investigations. Routledge, London

Hutto D (2008) First communions: mimetic sharing without theory of mind. In: Zlatev J, Racine T, Sinha C, Itkonen E (eds) The shared mind: perspectives on intersubjectivity. Benjamins, Amsterdam, pp 245–276

Ikegami T, Zlatev J (2007) From pre-representational cognition to language. In: Ziemke T, Zlatev J, Frank R (eds) Body, language and mind, vol 1. Embodiment. Mouton de Gruyter, Berlin, pp 197–240

Itkonen E (1978) Grammatical theory and metascience: a critical inquiry into the philosophical and methodological foundations of "autonomous" linguistics. Benjamins, Amsterdam

Itkonen E (2003) What is language? a study in the philosophy of linguistics. Turku University Press, Turku

Itkonen E (2005) Analogy as structure and process: approaches in linguistics, cognitive psychology and the philosophy of science. Benjamins, Amsterdam

Itkonen E (2008) The central role of normativity for language and linguistics. In: Zlatev J, Racine T, Sinha E, Itkonen E (eds) The shared mind: perspectives on intersubjectivity. Benjamins, Amsterdam, pp 279–306

Johnson M (1987) The body in the mind. University of Chicago Press, Chicago, IL

Johnson M (2005) The philosophical significance of image schemas. In: Hampe B (ed) From perception to meaning: image schemas in cognitive linguistics. Mouton de Gruyter, Berlin, pp 15–34

Johnson M, Lakoff G (2002) Why cognitive linguistics requires embodied realism. Cognitive Linguistics 13(3): 245–263

Johnson M, Rohrer T (2007) We are live creatures: Embodiment, American pragmatism and the cognitive organism. In: Ziemke T, Zlatev J, Frank R (eds.) Body, language and mind. Embodiment. Mouton de Gruyter, Berlin 1:17–54

Katz J (1981) Language and other abstract objects. Blackwell, Oxford

Kockelmans JJ (1967) Phenomenology: the philosophy of Edmund Husserl and its interpretation. Doubleday, Garden City, NY

Lakoff G, Johnson M (1980) Metaphors we live by. University of Chicago Press, Chicago, IL

Lakoff G (1987) Women, fire and dangerous things: what categories reveal about the mind. University of Chicago Press, Chicago

Lakoff G (1988) Cognitive semantics. In: Eco U, Santambrogio M, Violi P (eds) Meaning and mental representations. Indiana University Press, Bloomington, pp 119–154

Lakoff G, Johnson M (1999) Philosophy in the flesh: the embodied mind and its challange to western thought. Basic Books, New York

Langacker R (1986) Abstract motion. Proc Annu Meet Berkeley Linguist Soc 12:455–471

Langacker R (1987) Foundations of cognitive grammar, vol 1. Stanford University Press, Standord

Langacker R (2006) Subjectification, grammaticaliztion, and conceptual archetypes. In: Athanasiadou A, Canakis C, Cornillie B (eds) Subjectification: various paths to subjectivity. Mouton de Gruyter, Berlin, pp 17–40

Lotman J, Uspenskij B, Ivanov V, Toporov V, Pjatigorski A (1975) Thesis on the semiotic study of cultures (as applied to slavic texts). In: Sebeok T (ed) The tell-tale sign: A survey of semiotics. Peter de Ridder, Lisse, pp 57–84

Mandler J (1991) Prelinguistic primitives. Proceedings of the Annual Meeting of the Berkeley Linguistics Society 17:414–425

Mandler J (2004) The foundations of mind: origins of conceptual thought. Oxford University Press, Oxford

Meltzoff A, Moore MK (1983) Newborn infants imitate adult facial gestures. Child Dev 54:702–709

Merleau-Ponty M (1962 [1945]) Phenomenology of perception. Routledge, London

Montague R (1974) Formal philosophy: selected papers of richard montague. Yale University Press, New Haven, CT

Moran D (2005) Husserl: founder of phenomenology. Polity, Cambridge

Nerlich B, Clarke DD (2007) Cognitive linguistics and the history of linguistics. In: Geeraerts D, Cuyckens H (eds) The Oxford handbook in cognitive linguistics. Oxford University Press, Oxford, pp 599–607

Overgaard M (2004) On the naturalizing of phenomenology. Phenomenol Cogn Sci 3(4):365–379

Pateman T (1987) Language in mind and language in society. Clarendon, Oxford

Piaget J (1945) La formation du symbole chez l'enfant. Delachaux et Niestlé

Pinker S (1994) The language instinct. William Morrow, New York

Popova Y (2005) Image schemas and verbal synaesthesia. In: Hampe B (ed) From perception to meaning: image schemas in cognitive linguistics. Mouton de Gruyter, Berlin, pp 395–419

Putnam H (1981) Reason, truth and history. Cambridge University Press, Cambridge

Regier T (1996) The human semantic potential. MIT Press, Cambridge, MA

Rohrer T (2007a) The body in space: Dimensions of embodiment. In: Ziemke T, Zlatev J, Frank R (eds) Body, language and mind. Embodiment. Mouton de Gruyter, Berlin 1:339–378

Rohrer T (2007b) Embodiment and experientialism. In: Geeraert D, Cuyckens H (eds). The Oxford handbook of cognitive linguistics. Oxford University Press, Oxford, pp 25–47

Saeed JI (2003) Semantics, 2nd edn. Blackwell, London

Sartre J-P (1956 [1943]) Being and nothingness. a phenomenological essay on ontology. Washington Square, New York

Saury J-M (2004) The phenomenology of negation and its expression in natural language. Göteborg University, Department of Linguistics, Göteborg

Searle J (1983) Intentionality, an essay in the philosophy of mind. Cambridge University Press, Cambridge

Searle J (1995) The construction of social realiy. Penguin, London

Searle J (1992) The rediscovery of the mind. The MIT Press, Cambridge, MA

Searle J (2002) Consciusness and language. Cambridge University Press, Cambridge

Sheets-Johnstone M (2004) Preserving integrity against colonization. Phenomenology and the Cognitive Sciences 3(3):249–261

Sinha C (1999) Grounding, mapping and acts of meaning. In: Janssen T, Redeker G (eds) Cognitive linguistics: foundations, scope and methdology. Mouton de Gruyter, Berlin, pp 223–255

Sinha C, Rodríguez C (2008) Language and the signifying object: from convention to imagination. In: Zlatev J, Racine T, Sinha C, Itkonen E (eds) The shared mind: perspectives on intersubjectivity. Benjamins, Amsterdam/Philadelphia, pp 357–378

Sonesson G (1989) Pictorial concepts. inquiries into the semiotic heritage and its relevance for the analysis of the visual world. Aris/Lund University Press, Lund

Sonesson G (2004) Perspective from a semiotical perspective. In: Rossholm G (ed) Essays on fiction and perspective, Peter Lang, Bern, pp 255–292

Sonesson G (2006) The meaning of meaning in biology and cognitive science. Sign Syst Stud 34(1):135–214

Sonesson G (2007) From the meaning of embodiment to the embodiment of meaning: a study in phenomenological semiotics. In: Ziemke T, Zlatev J, Frank R (eds) Body, language and mind, vol 1: Embodiment. Mouton de Gruyter, Berlin, pp 85–128

Stern DN (2000 [1985]).The interpersonal world of the infant: a view from psychoanalysis and developmental psychology. Basic Books, New York

Talmy L (2000) Toward a cognitive semantics, vols 1 and 2. MIT Press, Cambridge, MA

Thomasson AL (2002) Instrospection and phenomenological method. Phenomenol Cogn Sci 2(3):239–254

Thompson E (2001) Empathy and consciousness. J Conscious Stud 8(5–7):1–32

Traugott E (1989) On the rise of epistemic meanings in English: an example of subjectification in semantic change. Language 65:ss. 31–55

Trevarthen C (1979) Communication and cooperation in early infancy: a description of primary intersubjectivity. In: Bullowa M (ed) Before speech. Cambridge University Press, Cambridge, pp 321–347

Turner M (2007) Conceptual integration. In: Geeraerts D, Cuyckens H (eds) The Oxford handbook of cognitive linguistics. Oxford University Press, Oxford, pp 377–393

Varela F (1996) Neurophenomenology: a methodological remedy for the hard problem. J Conscious Stud 3(4):330–350

Varela F, Thompson E, Rosch E (1991) The embodied mind. MIT Press, Cambridge, MA

Verhagen A (2005) Constructions of Intersubjectivity. Discourse, syntax and cognition. Oxford University Press, Oxford

Verhagen A (2008) Intersubjectivity and the architecture of the language system. In: Zlatev J, Racine T, Sinha C, Itkonen E (eds) The shared mind: perspectives on intersubjectivity. Benjamins, Amsterdam, pp 307–332

Vygotsky LS (1978) Mind in society: the development of higer psychological processes. Harvard University Press, Cambridge, MA

Wittgenstein L (1953) Philosophical investigations. Basil Blackwell, Oxford

Zahavi D (2001) Beyond empathy: phenomenological approaches to intersubjectivity. J Conscious Stud 8(5–7):151–167

Zahavi D (2003) Husserl's phenomenology. Stanford University Press, Stanford

Ziemke T, Zlatev J, Frank R (eds) (2007) Body, language and mind. vol 1: embodiment. Mouton de Gruyter, Berlin

Zinken J (2007) Discourse metaphors: the link between figurative and habitual analogies. Cogn Linguist 18(3):445–465

Zlatev J (1997) Situated embodiment: studies in the emergence of spatial meaning. Gotab, Stockholm

Zlatev J (2005) What's in a schema? Bodily mimesis and the grounding of language. In: Hampe B (ed) From perception to meaning: image schemas in cognitive linguistics. Mouton de Gruyter, Berlin, pp 313–343

Zlatev J (2007) Embodiment, language and mimesis. In: Ziemke T, Zlatev J, Frank R (eds) Body, language and mind, vol 1: Embodiment. Mouton de Gruyter, Berlin, pp 297–338

Zlatev J (2008) The dependence of language on consciousness. J Conscious Stud 15(6):34–62

Zlatev J, SEDSU-Project (2006) Stages in the evolution and development of sign use (SEDSU). In: Cangelosi A, Smith AD, Smith K (eds) The evolution of language: Proceedings of the 6th International Conference (EVOLANG6). World Scientific, New Jersey, pp 379–388

Zlatev J, Racine T, Sinha C, Itkonen E (eds) (2008) The shared mind: perspectives on intersubjectivity. Benjamins, Amsterdam

Zlatev J, Blomberg J, David C (in press) Translocation, language and the categorization of experience. In: Evans V, Chilton P (eds) *Language, cognition and space.* Equinox, London

Part VII
Applications and Expriments

The Role of Phenomenology in Psychophysics[1]

Steven Horst

Psychophysics is a branch of experimental psychology often described as being concerned with "the measurement of sensation". Some of the field's most important figures, like Gustav Fechner and S.S. Stevens, have viewed phenomenology – in the sense of the examination of the first-person experience of sensations and percepts – as playing a crucial role in psychophysics. But other practitioners and philosophers have been critical of this assumption. Some have held that what psychophysics really measures are functionally-characterized discriminative capacities. Others have taken the even more radical view that psychophysics does not really measure any inner variables, whether phenomenological or neural.

This article explores two sets of issues. The first is that of whether phenomenology plays a crucial methodological role in psychophysics, or whether psychophysical experiments can be understood solely in terms of measurement of discriminative capacities. I argue that Fechner and Stevens were correct in assigning to phenomenology a crucial methodological role in psychophysics, involving not only the report of properties of subjective intensities, but also simple intentional states such as seeing a figure *as* a square. This, I argue, cannot be understood merely in terms of discriminative capacities. And because psychophysics provides a substantial portion of the data which neuroscience and neural modeling must accommodate, the phenomenological properties that are essential to psychophysics are necessary precursors to much theoretical work in the modeling of processes underlying perception as well. As a result, some of the most rigorous and best-established areas of psychology cannot do without phenomenology

S. Horst
Department of Philosophy, Wesleyan University, Middletown, CT, USA

The second question arises from the radical challenge presented by writers who claim that psychophysics does not really underwrite the measurement of sensation, *or of any inner variable,* be it neural or phenomenological. The most important recent advocate of this view, Donald Laming (1997), holds that psychophysical experiments support only an *ordinal* scale for ranking stimulus intensities, and not the type of *ratio* scale that is needed to arrive at something like Fechner's Law, and thus count as a true *measurement* of sensation (or of anything else). The ratio metric, he goes on to claim, is an artifact of experimental and interpretive methods, and is furthermore not consonant with the full body of psychophysical data. Like me, Laming attributes an important role to intentionality in the actions performed by subjects in psychophysical experiments: specifically, he holds that subjects are not simply reporting on the intensities of sensations, but are *making judgments about properties of stimuli.* This introduces important complications into how we are to model the processes subjects are undergoing in psychophysical experiments. I shall argue that Laming's claims do not imperil realist assumptions about the phenomenology of sensation and perception, but do raise important concerns about the role phenomenology might play in psychophysical experiments, and about the "isomorphist" methodological canon urged by Todorovic (1987) and endorsed at a philosophical level in Horst (2005).

Fechner and the Birth of Psychophysics

Psychophysics was born out of the work of Gustav Fechner, particularly his *Elemente der Psychophysik* (Fechner 1860). Fechner's work, like that of his older contemporary Ernst Weber (1834), was concerned primarily with the measurement of sensations and the relation of the intensities of these sensations to those of the stimuli that occasion them. His measurements of sensation were obtained indirectly, through the method of just noticeable differences (JND). Fechner viewed his experiments as uncovering robust relationships between stimulus intensities and the intensities of subjective sensations. He claimed that these relations could be expressed by logarithmic laws of the form $S = \ln X$, with a different logarithmic base for each sensory modality.

While Fechner's work was concerned principally with the discovery of psychophysical laws expressing relations between the intensities of stimuli like light and sound and subjective experiential states, he also opined that the future of the science would eventually take a different turn, examining relationships between percepts and the physiological states in the nervous system that he assumed would be found to mediate the causal link between stimulus and percept. He distinguished these two types of projects as "outer" and "inner" psychophysics. Outer psychophysics is concerned with the relation between stimuli and percepts, while inner psychophysics is concerned with the relation between percepts and physiological states in the brain. The latter could not be directly measured in Fechner's day, but he believed that they would help to explain anomalies in the outer psychophysical data where the laws seemed to break down.

The word 'inner' in Fechner's 'inner psychophysics' seems principally to pick out what is *physiologically* "inner". (He understood his "outer" psychophysics, after all, to include percepts, which are "inner" in the alternative sense of being subjective experiences.) Generally, when Fechner spoke of inner psychophysics, he was speaking of the relations between percepts and the brain states that he presumed them to be specially related to. It is perhaps less clear whether Fechner understood inner psychophysics also to include the investigation of relations between stimuli and the physiological mechanisms underlying perception. Since there were no methods for examining neural states directly in his day, what he says about inner psychophysics is understandably more programmatic and less precise than what he says about outer psychophysics.

But as the relations between stimuli and brain states comprise an area of research that has developed significantly with the ability to measure neural activity, and is separable from the relation between that neural activity and experienced percepts, we would do well to give distinct names to (a) the study of relations between stimuli and neural states and (b) the study of relations between neural states and percepts. I shall thus divide what Fechner might have seen as the realm of inner psychophysics into two parts. The study of relations between stimuli and neural states I shall call *early psychophysics*, and the study of relations between neural states and percepts I shall call *late psychophysics*. The labels 'early' and 'late' are chosen in part to reflect Fechner's own non-materialist view, and allow more generally for the possibility that percepts are not identical with the neural states that play a role in their formation, but are separate events caused by neural events. However, I hope to do this without prejudice to the question of whether the relations of late psychophysics are best viewed as ones of identity, causation or some other relation. The relations in early psychophysics, by contrast, are clearly to be understood as causal relations (Fig. 1).

In this schema, we may follow Fechner, with some modifications, in viewing outer psychophysical laws as expressing causal relations between stimuli and percepts which we may decompose into two stages. The first ("early") stage consists in causal relationships between stimuli and neural processes in the brain. The second ("late") stage consists in relationships of some sort (e.g., causation or identity) between these neural processes and experienced percepts.

The word 'psychophysical' (sometimes rendered in hyphenated form, as 'psycho-physical') is also used by philosophers of mind, generally to express relations between experiential or mental states and their "neural correlates". Philosophers, however, tend to use this term slightly differently from psychophysicists. For psychophysicists, the relations called "psychophysical" tend to be lawlike relations between *types* of states. Philosophers, by contrast, tend to apply the word to relationships between *token* states (e.g., assertions of token identity), even when laws and measurements are not directly involved. Some, such as Davidson (1970), go so far as to deny that there *are* any psychophysical *laws*. Closer attention to Davidson's arguments, however, reveals that the "psychological" side of "psycho-physical" relations he is concerned with consists, not in the sensory states that are studied by scientific psychophysics, but in intentional states like beliefs and desires. His arguments

Outer Psychophysics

Inner Psychophysics
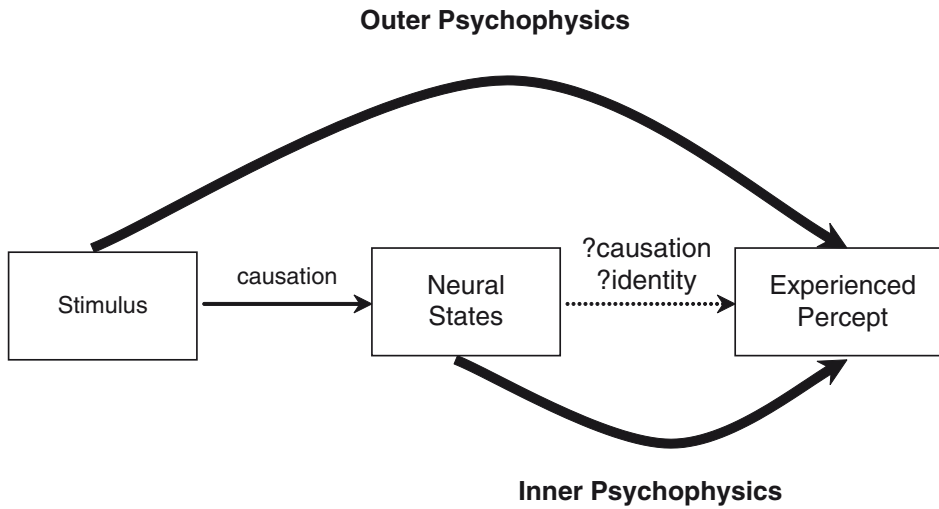
Fig. 1 *Fechner's Inner and Outer Psychophysics*. Fechner's outer psychophysics was to be concerned with relations between stimuli and percepts. However, he supposed that this relation was mediated by neural processes which were either identical with or a cause of the subjective percept. The relation between the neural state and the percept was to be treated by his inner psychophysics

Outer Psychophysics

Early Psychophysics   **Late Psychophysics**
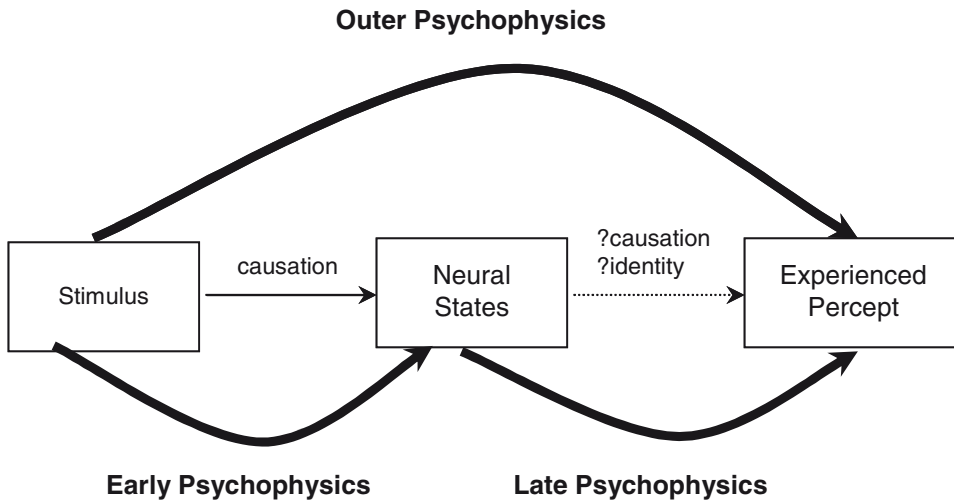
Fig. 2 *Fechner's Model Expanded*. This model makes the two stages Fechner supposed to underlie outer physics more explicit, giving a distinct name to the study of each. Early psychophysics studies the relations between stimuli and neural states, late psychophysics the relations between those neural states and subjective percepts

depend upon assumptions about the indeterminacy of translation and an interpretationist semantics that apply only to such intentional states, and these assumptions are not readily transportable to considerations about sensations and percepts. While I do believe that intentional states often have an inherent phenomenology, and that some

states that figure in psychophysical data involve intentionality, I regard the particular "psycho-physical" relationships towards which Davidson's arguments are directed as altogether different in kind from those of scientific psychophysics, and so shall not consider them further here. Instead, I shall confine myself to philosophical considerations about the relation of phenomenology to psychophysics in Fechner's sense.

Developments Since Fechner

While Fechner's work drew immediate criticism on a number of fronts, psychophysics has flourished as a scientific discipline. Indeed, it was one of the first of the cognitive sciences to achieve mathematical exactitude, often viewed as one of the hallmarks of scientific maturity; and it has supplied much of the data for theories of the mechanisms underlying perception, from Helmholtz's three-color theory of vision to contemporary work in the neurophysiology of perception and computational neuroscience. A number of aspects of Fechner's own work, however, have been either rejected or called into question in subsequent generations.

Like his contemporary Wilhelm Wundt, Fechner believed that subjective states comprise a major and indispensable portion of the data of the psychological sciences. While he hoped that inner psychophysics would eventually supersede outer psychophysics, he envisaged experiential states as comprising one class of the relata in both disciplines. This view was retained by some of the most influential psychophysicists of the twentieth century, particularly S.S. Stevens, who went even further than Fechner in this direction, holding that it is possible for subjects to make methodologically valid quantitative judgments about the intensities of percepts, such as that one is twice as intense as another, thus obviating the need for the JND method. On the other hand, many other psychophysicists held to Fechner's experimental method or refined it, while rejecting his interpretation of it. They held that the JND method does not measure subjective intensities at all, even indirectly, but rather measures *discriminative abilities*. Some behaviorist psychologists went even further, rejecting both types of relata involved in inner psychophysics – subjective experiences and neural states – on methodological grounds.

Other psychophysicists called into question either Fechner's own interpretation of his data in the form of *logarithmic laws* or even the broader claim that the data really show lawlike phenomena at all. Most famously, Stevens argued that the data are better described by a power function than by Fechner's logarithmic laws. Donald Laming (1997) has gone even further, claiming that the data support only an ordinal ranking of intensities, and not any *measurement* of sensation in terms of algebraic relations.

The domain of psychophysics has spread from measurements of intensities of stimuli and sensations to large classes of sensory effects. Much important work in visual psychophysics, for example, has been devoted to visual illusions, such as subjective contour figures, the Cornsweet effect, the Hermann grid, and the Ehrenstein effect. Such work is still concerned with relations between stimuli and

percepts. But it is not limited to the measurement of individual sensations, as it includes gestalt phenomena as well.

What Does Outer Psychophysics Measure?

While contemporary psychophysics may be seen as a descendent of Fechner's work, subsequent generations of psychophysicists have framed a direct challenge to Fechner's *understanding* of psychophysics. Fechner and Stevens both held that some of the relata of outer and inner psychophysics are *percepts*, understood as experiential states. Fechner was aware of difficulties in measuring percepts. (As, of course, was Stevens, writing almost a century later.) Unlike Stevens, who believed that subjects could directly assign a scale of intensities to percepts, Fechner approached the problem of measuring them indirectly, through the JND method, which more directly measures what pairs of stimuli subjects can differentiate. (For example, to correctly identify one as more intense than another, or as differing in other qualities like hue.) But this leads to a possible problem for Fechner's interpretation: *why interpret such experiments as measuring subjective intensities at all, as opposed to merely measuring discriminative abilities?*

We might look at the issue with the use of a diagram. Fechner assumed that stimuli cause neural states, which in turn either *cause* or *are identical with* sensations. He took subjects to be reporting on perceived differences between stimuli on the basis of their sensations. One interpretation of this would be that the basis of their reports was a comparison of sensations through introspection of their subjective intensities (Fig. 3a). But one might equally well suppose that the basis of the discrimination lies in a mechanism that compares neural states. Neural states might thus both give rise to sensations and, through a separate process, provide the basis for discriminative abilities. (Figure 3b) Or indeed, equally consistent with the data, the neural states might provide the basis for discriminative abilities without subjective sensory appearances playing any role in the story at all (Fig. 3c). Given that the neural circuitry is to play the crucial explanatory role in inner psychophysics, the phenomenology of the sensation seems to be left with no role to play, and can either be eliminated as an unnecessary theoretical posit or at least safely ignored for purposes of science.

Now that we have the ability to measure neural activity directly, some psychophysicists have taken a different and more radical route, dispensing with the profile of the percept as reported even as a constraint upon successful modeling. Consider the following case, involving the Cornsweet effect (see Fig. 4). This illusion (4a) is a visual effect in which two adjacent areas appear to each be of uniform internal brightness, with one darker than the other, thus having a step-shaped percept profile (4b). (The subjective intensity of the percept is referred to as "brightness" and the objective intensity of the stimulus as "luminance".)

One way of inducing such a percept, of course, is to present a stimulus with a similarly step-shaped luminance profile (4c). But it is also possible to induce the same
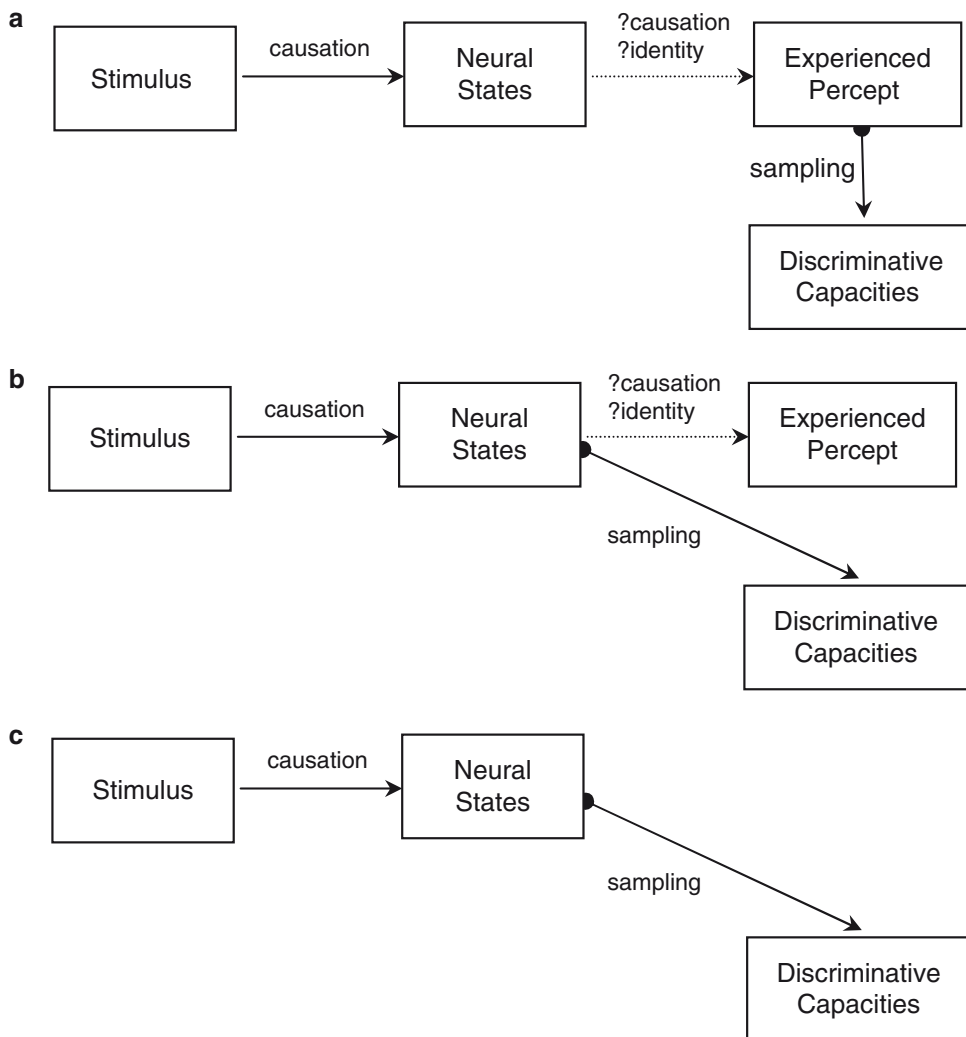
Fig. 3 *Three Models of the Psychophysical Process.* The first diagram (**a**) represents a process in which a stimulus causes a neural process which in turn either causes or is identical with a subjective percept. The subject then makes discriminations based on introspective sampling of this percept. In (**b**), the neural state gives rise to discriminative abilities through a separate causal pathway, in which the pre-existing neural state is sampled by some mechanism that results in a discrimination. In (**c**), the basis of discriminative abilities is as in (**b**), but the subjective percept is omitted from the model as an explanatory dangler

percept by other means, such as a luminance profile with two ramps (4e), or with a cusp at the boundary (4e). When the stimulus has a ramp-shaped profile, each subregion has an internal change in luminance that is not perceived, but they have identical luminance profiles. When the stimulus profile is cusp-shaped, the luminance profiles of the subregions are identical except for the regions surrounding the cusp.

Some theorists (Cornsweet 1970; Campbell et al. 1978; Ratliff 1978; and Ratliff and Sirovich 1978) have suggested that the production of the effect is

Illusion

Brightness Gradient of Percept
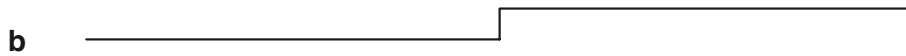
Luminance Gradients of Inducing Stimuli
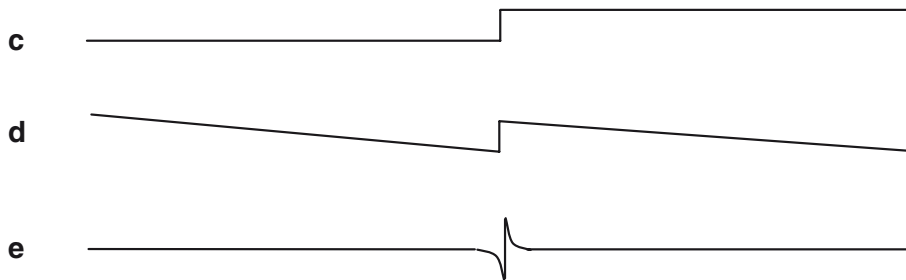
Fig. 4 *The Cornsweet Illusion*. The illusory figure (**a**) appears to be composed of two squares, each of a consistent internal brightness, but with the right figure brighter than the left. This subjective experience of the percept is represented by the step-shaped luminance profile (**b**). Such a percept can, however, be induced in a variety of ways. It can be induced by a stimulus with a step-shaped luminance gradient (**c**), in which case it is not illusory. But it can also be induced by a ramp-shaped luminance gradient (**d**) or a luminance gradient with a cusp at the boundaries (**e**). (The actual figure presented has a luminance gradient produced by ramped luminance. The squares both get brighter internally from left to right, as can be verified by occluding the boundary.)

explained by the fact that the stimuli all share features in terms of Fourier analysis that can produce equivalent effects in known neural models. As Todorovic (1987) summarizes it,

> In terms of Fourier analysis, the two distributions have similar high-frequency content but different low frequency components. However, the visual system is relatively insensitive to low-spatial-frequency stimulation (Campbell and Robson 1968). According to Cornsweet (1970), Campbell et al. (1978), Ratliff (1978), and Ratliff and Sirovich (1978), these facts

COCE - Isormorphic and Non-Isomorphic Theories
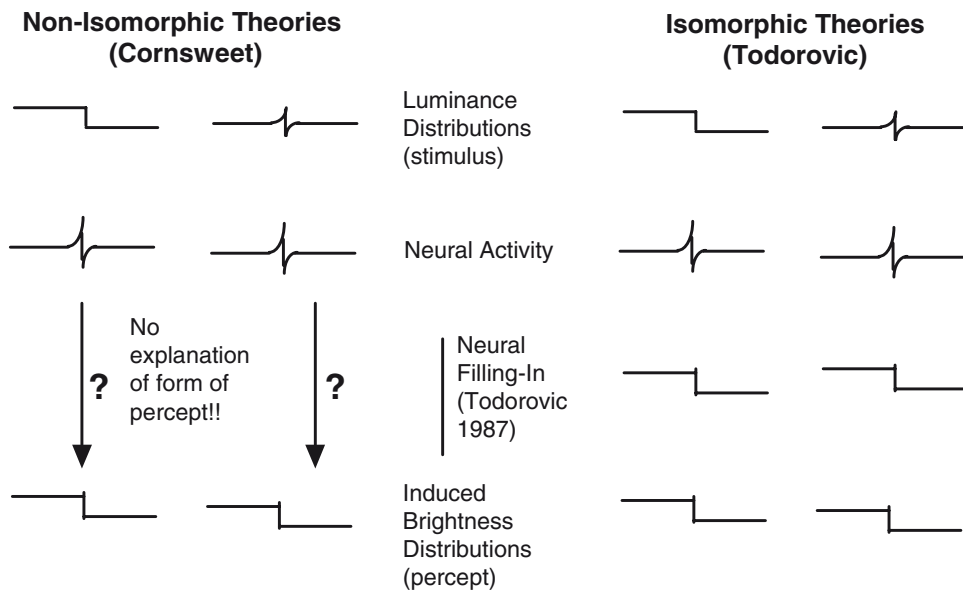
Fig. 5 *Isomorphist and non-isomorphist explanations of the Cornsweet Illusion*. Luminance patterns with steps and luminance patterns with cusps both produce similar neural activity at some stage in early vision and similar brightness percepts. However, the profile of the neural activity does not match (is not isomorphic to) that of the percept. Non-isomorphist theories provide no explanation of how a cusp-shaped neural profile yields a step-shaped brightness profile. Isomorphist theorists like Todorovic (1987) insist that explanation is not complete until a neural correlate is found whose activity profile matches that of the percept. Diagram based on Todorovic (1987), p. 546

> amount to an explanation of the (Cornsweet effect). The cusp-shaped and step-shaped distributions look similar because their effects are similar: the visual system suppresses the aspects of these stimuli that differ (shallow spatial variation of luminance), and transmits more faithfully the attribute they have in common (abrupt change) (p. 547).

Todorovic complains, however, that the profile thus produced does not reflect the step-shaped contour of the percept. To model the production of such a profile, the model requires an additional level of activity, which he describes as "filling-in".

Todorovic is here raising important methodological concerns: namely, *When is explanation at an end?* and *Just what is in need of explanation*? For Cornsweet and others, what seemed in need of explanation was simply the fact that the visual system *treats certain distinct classes of luminance profiles as equivalent*, and explanation is complete when one identifies a plausible candidate neural circuit that treats the different types of stimuli identically. For Todorovic, this is not enough. What is in need of explanation is the *particular brightness profile* of the percept, and explanation is not complete until one has a model whose output has a profile matching that of the percept (e.g., step-shaped rather than cusp-shaped). Todorovic distinguishes

between these approaches as "non-isomorphist" and "isomorphist", respectively, the names reflecting theorists' views on whether it is necessary for a neural model to produce a profile isomorphic to that of the percept.

What is at stake here, philosophically? The isomorphist approach to modeling accords the percept a privileged role in constituting the data that visual modeling needs to explain. Matching the percept is a constraint upon successful modeling. The non-isomorphist approach does not accord this role to the percept. Instead, modeling is complete when there is a level of neural explanation that accounts for the fact that two sets of stimuli are indistinguishable to the perceptual system.

It seems clear, however, that (all else being equal) one ought to prefer the isomorphist approach. A model that explains why two stimuli are indistinguishable from one another, but not why they look the particular way they do, has surely left *something* out. There is a very reasonable question, for example, as to why *cusp*-shaped neural activity would result in a *step*-shaped percept. If we assume that, in human perception, a step-shaped percept *is* produced, then either the neural models that fail to answer this question are incomplete, and leave out a further stage at which step-shaped neural activity is produced, or else there is an abiding mystery here as to why there is a mismatch between the neural and the phenomenological properties. At least to the physicalist, there should be an intuition to the effect that such gaps ought to be regarded as further research agenda, and not as mysteries. But this means that even the physicalist needs to accord to phenomenology – at least the sort that is involved in the reporting of such percepts – a privileged and ineliminable role in the methodology of perceptual modeling. The percept is what supplies crucial data that are in need of explanation; and without it, we do not know when explanation is properly at an end. Non-physicalists, of course, might be more willing to countenance the possibility that neuroscience might find *no* level at which neural activity is isomorphic to the profile of the percept. But she is committed to the phenomenology of the percept on other grounds, as a distinctive type of feature of perceptual experiences as such.

Not Just Intensities Anymore

A great number of important sensory effects measured by psychophysicists involve more than simple measurement of stimulus *intensities*. Consider, for example, subjective contour figures of the sort shown in Fig. 6. In such figures, one "sees" figures – e.g., the square in this example – that are "not really there", in the sense that there are no objective lines or luminosity gradients located where we "see" the edges of the figures. Moreover, we also experience an illusory contrast in brightness: the figure looks brighter than the background, even though it is of equal luminosity.

Subjective contour figures are standard textbook examples of psychophysical effects, but they involve something more than a simple comparison of intensities:
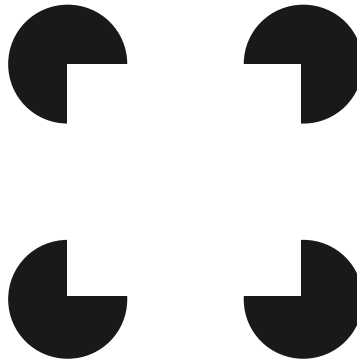
Fig. 6 *Subjective Contours*. This figure, called the Kanizsa Square, is an example of a subjective contour figure. The perceiver should "see" the (non-existent) boundaries of the square filled in, and experience the interior of the illusory square as being slightly brighter than the surround

they involve a gestalt phenomenon in which one region of the stimulus is constituted as an object. In a model of the processes that produce these illusions (as well as normal vision), one would want, at very least, to have an output level at which the values corresponding to perceived brightness for the figure are greater than those for the ground. But arguably one would also need something more, and radically different in kind: namely, (a) something corresponding to the separation of figure from ground, and (b) something corresponding to the "invisible edges" that the mind somehow "sees" even though there is, in a perfectly straightforward sense, no such thing there to be seen. (Hence the illusory character of the percept.) There are models that attempt to do these things, though the assessment of their adequacy would lie outside the intended scope of this essay. (Cf. Grossberg and Mingolla 1985a,b, Grossberg 1987, Grossberg and Todorovic 1988, Grossberg 1994.)

Rather, I wish to point out something that is interesting and perhaps surprising about these examples from a philosophical standpoint. This is the fact that, as the perception of illusory contour figures involves seeing something *as* a figure of a particular kind, it involves at least a primitive sort of *intentionality*. (Brentano 1874) That is, in the inclusion of such effects, psychophysics has crossed over the divide that philosophers and psychologists alike have often drawn between sensation and intentionality. The percept in such cases *is* something like what we would express by saying "a square that is brighter than its background, with darker circles partially occluded behind the corners". It is not clear to me that there is *any* way to capture the features of the percept *without* utilizing intentional notions like "seeing-as-a-square". The kind of "seeing" involved in such cases is the type that is expressed in the intentional idiom, rather than a shorthand for a mere conglomeration of sensory states (individual qualia) across the visual field. Indeed, as these particular effects involve *illusory* stimuli, reports of them are clear cases of intentional contexts. (Chisholm 1957) That is, a subject viewing a subjective contour figure can honestly say "I see a square" even if there is no objective square there to be seen. Indeed, this implication would seem to ramify to all illusions, regardless of modality, that are properly reported with an intentional verb and an embedded clause reporting on the content of the perception.

Why is this interesting? One reason is scientific, and has to do with what such examples show about the mind. One thing that subjective contour figures seem to suggest is that the things we call "sensation" and "cognition" are in fact inter-connected, even at the fairly low level of visual processing involved in figure-ground separation. When the mind or brain constitutes something as an object, this can affect "sensory" qualities like perceived brightness as well.

A second reason is methodological and philosophical. Put simply, not only subjective intensities, but also certain types of intentional notions, seem ineliminably to be part of the data of psychophysics. This strikes me as crucially important for discussions about intentional states that treat them as "theoretical posits" of "folk psychology". This characterization of intentional states can work only if the data of psychology and psychophysics can be captured without recourse to the intentional idiom. That this is problematic when considering the data of cognitive psychology is already well-attested in decades of criticisms of behaviorism. But one might have thought that at least perceptual psychophysics might prove more amenable to this approach. However, this seems not to be the case after all. Among the effects that constitute the body of data supplied by psychophysics for projects of perceptual and neural modeling, some at least of them seem to be ineliminably couched in the intentional idiom, and indeed to involve simple, but garden-variety, intentional states. And since standard arguments for eliminativism apply only to true theoretical posits, those arguments cannot get off the ground in these cases. (Compare Horst 2005b.)

A More Radical Concern: Laming's Challenge

A far more radical issue has been raised by a number of writers over the years (Cf. von Kries (1882) and James (1890), and more recently taken up by Savage (1970), Zuriff (1972), Tumarkin (1981), Boynton (1989), Laming (1997).) Their claim is that the data of various psychophysical experiments do not in fact reflect measurements, either of sensation *or of the intensity of any inner variable*. The most recent systematic discussion of this view is that of Donald Laming, and it is with his treatment of the issues that I shall principally concern myself here. Laming is himself an experimental psychologist and not a philosopher, and the issues he pursues have relatively little contact with the issues that philosophers of mind or of psychology have traditionally studied. However, for philosophers of psychology, at least, it is important to pay attention to the methodological issues that arise for the practitioner. These are both important in their own right and may also have broader philosophical implications.

We have already discussed the difference between the methodologies pursued by Stevens and by Fechner. The majority of psychophysicists have followed Fechner, and have criticized Stevens's methodology of having subjects directly assign intensity values to their percepts. On the other hand, Stevens's mathematical characterization of the intensity laws as power functions has tended to attract more adherents than Fechner's characterization in terms of logarithmic functions. Laming explores

these issues as ably as anyone I have read. However, my concern here is with two more distinctive and radical claims that he makes about the psychophysical data. Laming first points out that the standard statements of psychophysical laws found in Fechner or Stevens really involve multiple claims that can be separated from one another. Most importantly, that:

1. The experiments performed by psychophysicists provide an adequate basis for assigning the type of metric to *whatever* it is that is correlated with stimulus intensity that is needed to underwrite an algebraic law (i.e., a ratio metric).
2. This amounts to a measurement of relations between stimulus intensity and some feature of *sensations*.

Some writers he surveys (e.g., Stevens) seem to make different assertions at different times, sometimes asserting a version of (1) that links stimulus intensity to some *neural* variable, and at other times linking it to *sensation*. Laming argues that *both* of these claims are false. We have already examined one type of attack on the second claim, from those who view psychophysics as measuring only discrimination capacities, and I have given *prima facie* reason to reject such attacks. Laming, however, devotes most of his book to the first assumption, which arguably strikes more at the heart of psychophysics. And if psychophysics does not measure *anything*, it follows trivially that it does not measure sensation. It is impossible to do full justice to all of the details of Laming's argumentation here, but I shall attempt to summarize key points in his argument.

The nub of the matter is that the type of "measurement" required to underwrite the kinds of algebraic laws put forward by Fechner and Stevens requires that data points be ordered on a particular type of scale, a *ratio scale*. There are also weaker types of scales, most importantly, *ordinal* scales, which arrange values comparatively (e.g., as greater or less than one another) but do not license addition or division. Both Fechner's logarithmic law and Stevens's power law require a domain that can be ordered according to a ratio scale. One of Laming's central claims is that the psychophysical data support only an ordinal scale and not a ratio scale, and hence do not really license the kind of algebraic interpretations found in Fechner's or Stevens's laws.

One important type of evidence Laming marshals comes from comparing the data from different types of tests, such as those requiring difference-detection versus those that require increment detection or cross-modal matching. Fechner was able to arrive at a logarithmic law because his tests involved only difference-detection, and because he made additional assumptions, such as that all JNDs are subjectively equal. The crucial role played by such additional assumptions suggests that Fechner's algebraic interpretation of the data points may be an artifact of his methodology. And this suspicion seems to be supported by the fact that data collected through other means, such as increment detection, cannot be reconciled with those arrived at through Fechner's method through a single ratio scale. Similar considerations apply to Stevens's method. As Laming summarizes the argument:

1. If numerical magnitude estimates establish a valid scale of sensation, then that scale must be a *ratio scale*, because 'equal stimulus ratios tend to produce equal

sensation ratios' (Stevens 1957[b], p. 162), and the relation of sensation to the physical magnitude of the stimulus must be a power law.

2. If that ratio scale is extensive, not only are ratios of sensation meaningful, but *differences* as well, because an extensive ratio scale also has interval scale properties.

3. It then follows that magnitude estimates of differences in sensation should conform to the same power law as estimates of ratios. This is contrary to experimental observation. (1997)

Perhaps more surprisingly, Laming claims that the very same problems arise in trying to match stimulus intensities with intensities of *neural activity* through direct sampling, and hence they cannot be avoided by retreating from outer to early psychophysics. The problem, if you will, is not so much that psychophysics does not measure *sensation* (as opposed to something else), but that it does not *measure* sensation (or anything else).

Laming's alternative interpretation of the data involves several separate theses. The first of these is that what experimentalists like Fechner took to be reports of subjective intensities were in fact reports of perceived properties of the stimuli (in the sense of "properties-of-stimuli-as-perceived"). Previous writers, he opines, tended to fall into what Boring (1921) called "stimulus error" (see also Lockhead 1992). In perception, we are in fact making judgments about the intensities of *properties of objects*. The stimulus error consists in mistaking these judged-intensities-of-properties-of-objects for intensities of sensations. (This is more or less the converse of the *attribution error*, of projecting properties of one's psychological states onto external objects towards which they are directed.)

Laming's second claim is that issues with the variances in psychophysical data obtained from different subjects and under different conditions suggest that subjects' discrimination abilities are consistent with no better than an ordinal scale. Actually, it is a somewhat complex ordinal scale, with five values: *much less, less, equal, greater* and *much greater*, rather than the two-valued *equal/unequal* or the three-valued *less/equal/greater*. However, the crucial point is that it is not a scale that supports algebraic operations of addition or division.

Third, when subjects are required to give numerical estimates of differences or ratios between stimuli, they are in fact engaged in a process of *judgment*, which does not simply reproduce a scale already implicit in pre-existing neural or phenomenological states, but which produces artifacts of its own. One possibility he explores for the origins of this imposed scale involves research on the *subjective values of numbers:* briefly glossed, human behavior in making numerical comparisons of stimuli actually results in results with a logarithmic bias, rather than conforming to canonical arithmetic. (One way of understanding this would be that humans have multiple internal modules for different types of mathematical operations, some of which conform to canonical arithmetic laws little better than ordinary reasoning conforms to Bayesian rationality.) Laming supports his thesis by showing that his five-value ordinal scale can help account for various data obtained by different experimental methods that proved recalcitrant so long as one assumed

there must be a single ratio scale of internal intensities that is measured by psychophysical experiments.

I shall not explore the details of Laming's argumentation that the data points in intensity experiments do not support more than an ordinal metric. Instead, let us ask *what follows if Laming is correct here*. The obvious philosophical questions are those of (1) whether we are still entitled to speak of psychophysical *laws* if the data do not support an algebraic interpretation such as those of logarithmic or power functions, and (2) if they are not laws in this sense, what implications this might have for the status of psychophysics? And to do this, it is necessary to bring into the discussion some broader considerations from recent philosophy of science.

The first question is to some degree a matter of semantics. Philosophers have often used the term 'law' very broadly – for example, in a fashion that could apply to universally quantified propositions like "all swans are white." In physics, by contrast, the word 'law' tends to be reserved for things expressed in algebraic equations. However, we also speak of chemical "laws" that capture what typically happens when compounds are combined, and these are dissimilar in form to the laws found in, say, theory of gravitation or thermodynamics, so the usage is not completely univocal in the sciences themselves, even if we confine ourselves to the physical sciences. Moreover, there are types of formally-exact modeling that need not be in the form of algebraic equations, but might, instead, be in the form of a computer program. And if there is a well-defined ordinal scale that unifies a given body of psychophysical data, this can be rendered formally exact even if it does not allow the application of algebraic computations. Indeed, Laming's own attempts to explain data from different types of experiments by use of his five-value ordinal model only lend support to this point. In short, Laming's arguments support the conclusion that psychophysical data do not support "laws" in a particular narrow sense, but not that they do not support any type of mathematically-rigorous modeling. And in this regard, they are in no worse shape than the generalizations of many other sciences.

What about the concern that the laws put forward by Fechner and Stevens are unable to accommodate all of the relevant data? At this level, at least, it looks as though Laming is pursuing a very ordinary line of argumentation that is commonplace in the sciences: that if theory A accounts for everything covered by theory B, plus things B cannot account for, we should prefer theory A. And it may well be that Laming's account is in fact to be preferred on these grounds. But this is not the issue I wish to address. My question is, rather, *if* this is so, what should we conclude about the status of Fechner's (or Stevens's) Law? And the answer to that is not so simple.

As Nancy Cartwright (1983) has shown us, even fundamental physical laws like the gravitation law do not license exact predictions of real-world behavior. Indeed, precisely because they isolate real variables (like gravitational force) singly, they screen out other variables (like magnetism and wind resistance) that tend to be at work in real-world situations. Their *theoretical power* is bought through idealizing away from other factors that influence real-world behavior, and hence they fail to predict the actual data one would obtain either in the laboratory or *in vivo*. So the mere fact that a law yields only a rough fit with laboratory data does not, in itself, mean that there is something wrong with the law.

462 S. Horst

It might get something *right*, but that something may not be something we can isolate experimentally in the laboratory.

There is, of course, an important difference between the neural processes underwriting the psychophysical transformations and laws of basic physics, in that the latter are fundamental. Whatever else we may think about perceptual processes, we must suppose that they are at least partly emergent out of complex neural processes, involving complicated feedback processing in multiple cortical and extracortical areas, and dependent upon lower-level variables such as transmitter levels and the precise topology of neural connections that make up any particular brain. A model of gravitation excludes electromagnetism, but gravity and electromagnetism are independent forces. A model of one area of the brain likewise excludes the activity of other areas, but those areas may actually play a role in modulating the behavior of the area modeled. And a model at the level of a neural circuit also idealizes away from variations in lower-level processes, such as variations in the levels of various neurotransmitters, which in real life affect the behavior of the circuit. (By contrast, with fundamental physical forces, there presumably are no lower levels to take into account.) The modeling of neural systems is thus, in a principled way, more complicated than models in basic physics (see Horst 2007).

This, in turn, points to an important possibility: that generalizations like Fechner's Law may in fact be very good models of *some component* of the perceptual process without being good models of the *entire* process. An example might help to make this clearer. In the study of color vision, Helmholtz made a seminal contribution in proposing the three-color model. This accounted for an important body of laboratory data, but left other data unaccounted for, such as the fact that there is a phenomenologically-pure yellow but not a phenomenologically-pure orange. The latter phenomenon was successfully modeled by Hering's color opponency theory. Helmholtz's model was, in one sense, not a "good" model of color vision, in that there are phenomena of color vision it failed to account for. But it turned out that it did a much better job of describing one component of the processing of visual information: the behavior of three types of cone cells in the retinas of trichromats. (Hering's model likewise found a plausible neural localization in the operation of retinal ganglion cells.) Had we simply rejected Helmholtz's model and sought one that was a better mathematical fit with the data, without regard to how the process was performed by neural mechanisms, we might have missed out on the crucial insight that color vision involves multiple stages of processing, one of which is in fact rather well described by the three-color model.

The upshot of this, in my view, is that it is a mistake to try to evaluate interpretations of psychophysical data in isolation from models of their neural implementation. It may turn out that the logarithmic form of Fechner's own data is a pure artifact of the experimental set-up or of additional assumptions Fechner made in interpreting his data. But, alternatively, it may turn out that that same, rather artificial, set-up has succeeded in isolating the formal shape of one or more components of the perceptual process by minimizing the contributions of other components. Of course, it may also turn out that the "component" in question comes into play only when the subject is asked to do certain things, like make comparative judgments

involving the imposition of arithmetic values. My point is not to advocate one interpretation over another, but to urge that we are now at a point where our interpretation of psychophysics can and should be carried out in tandem with other enterprises, such as computational neuroscience. Given that sensation and perception are products of complex neural processing, we probably should not assume that all outer psychophysical data should be accommodated in a model as simple as an algebraic law. Rather, we should expect that their modeling would require something like a circuit diagram of the neural processes involved, along with the mathematical laws that govern *them*. Such a model would still allow us to speak of "psychophysical functions". They would just be much more complicated functions that cannot be expressed in simple equations like Fechner's Law.

Indeed, while Laming's book does treat psychophysics more or less in isolation, I think that some of his claims, such as the role of judgment in the shaping of some of the psychophysical data, ought themselves to lead us to seek models that treat perception and sensation as complex neural processes. And his suggestions here *do*, I think, lead to some concerns about the "obvious" interpretations of psychophysical data and laws.

Let me attempt to reconstruct such a concern as a philosopher, in a fashion that I hope will not do injury to Laming's own understanding of it. In our earlier diagrams, we represented the *report* given by an experimental subject as a simple reproduction of something already present in mind or brain: either a phenomenological state or a neural state. But in fact, this need not be the case. A subject who is simply presented with a stimulus, and a subject presented with a stimulus *who is also making a judgment about it*, are not in exactly the same mental or neural state. Even if we assume that the both subjects' states share a common element, one is doing something that the other is not. Subjects in psychophysical experiments are asked to *report*, and reports can both be *inaccurate* and introduce *artifacts* due to whatever processes go into sampling a pre-existing state and formulating a verbal response. One might hope that the inaccuracies would not be systematic and would thus be corrected by an adequately large sample. (Though this by no means need be the case.) More serious, to my mind, is the concern that the process of sampling and reporting *adds something* that was not there before. Some aspects of this, like artifacts introduced by translating sensory states into verbal reports, might be avoided through experimental methods that do not require verbal reporting. But to the extent that the subject is posed with a particular cognitive task, such as comparing two stimuli (whether algebraically as per Stevens or as same/different as per Fechner), it is possible (a) that the very state examined is altered by the sampling process, and (b) that the direct basis for the report is *not* a phenomenological or a sensory-neural state, but some other representation produced by the act of sampling and comparison.

I think it is useful to go beyond Laming here in exploring several additional models of what may go on in between stimulus and report. Our previous models of the processes underlying the data of outer psychophysics assumed that *either* neural states involved in ordinary sensation *or* subjective percepts (or both) would be characterized by properties that were faithfully reproduced in the reports from which data were obtained. Generalizing from our earlier models, we may represent

Stimulus Profile   Percept Profile   Verbal Report

Neural Profile
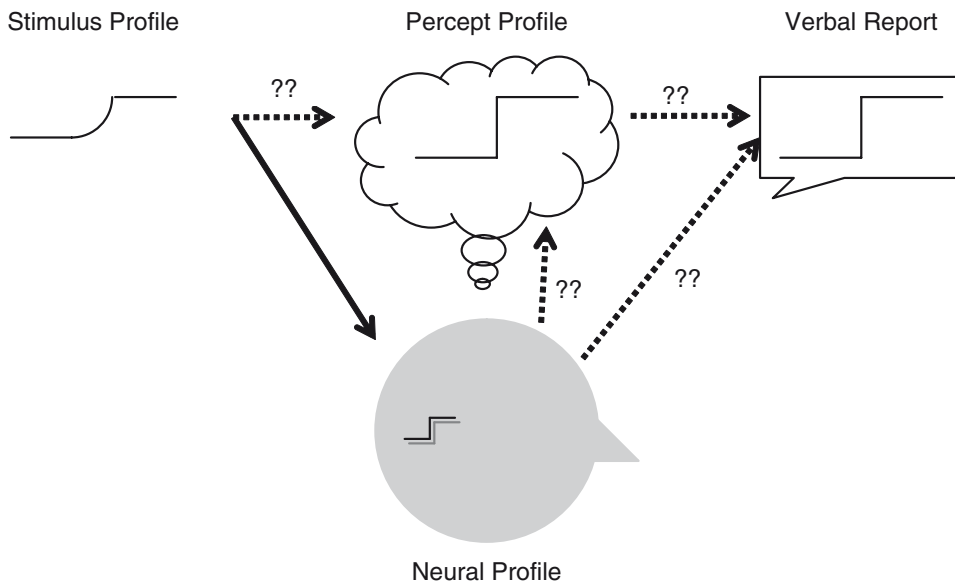
Fig. 7 *Meta-Model of Psychophysical Relations*. This diagram compresses the alternative models in Fig. 3. Broken lines indicate possible relations of causation and/or identity to accommodate the various possibilities. Note that all variants on the model assume that there is a subjective percept and/or a neural state whose profile is faithfully reproduced in the report

this with Fig. 7. Here many of the causal arrows are shown with dotted lines in order to accommodate variations on this model that differ with respect to how they take the ultimate production of a report to be produced. The neural profile in question, importantly, is supposed to be one that is present when the subject is presented with the stimulus, *regardless of whether she is also asked to attach a value to it or compare it with something else*.

But if Laming is correct, there is something wrong with this meta-model, and with all its variations: namely, that the subjective report is not simply a faithful reproduction of the properties of a pre-existing neural or phenomenological event, but is a product of an additional step introduced by the experiment. How ought we to represent the process of sampling or comparison? One possibility, or family of possibilities, would require that we postulate an additional mental/neural module that is involved in *representing* either the phenomenological or the neural state. It is cued by the subject who has received experimental directions, and takes as input, let us assume, a "raw" sensory or state with its own activation profile. Thus in Fig. 8, we now have representations of not one but two neural states, one representing the "raw" sensory trace, which is presumably present whether the subject is asked to compare, quantify or otherwise probe her reactions, and a second representing the results of such a probe. It is the latter which corresponds to the experimental data.

In Fig. 8, I have drawn these two neural states with different profiles, though of course it is an open question whether every such sampling process in fact distorts the "raw" state. The important thing to note is that it *might* misrepresent it, either
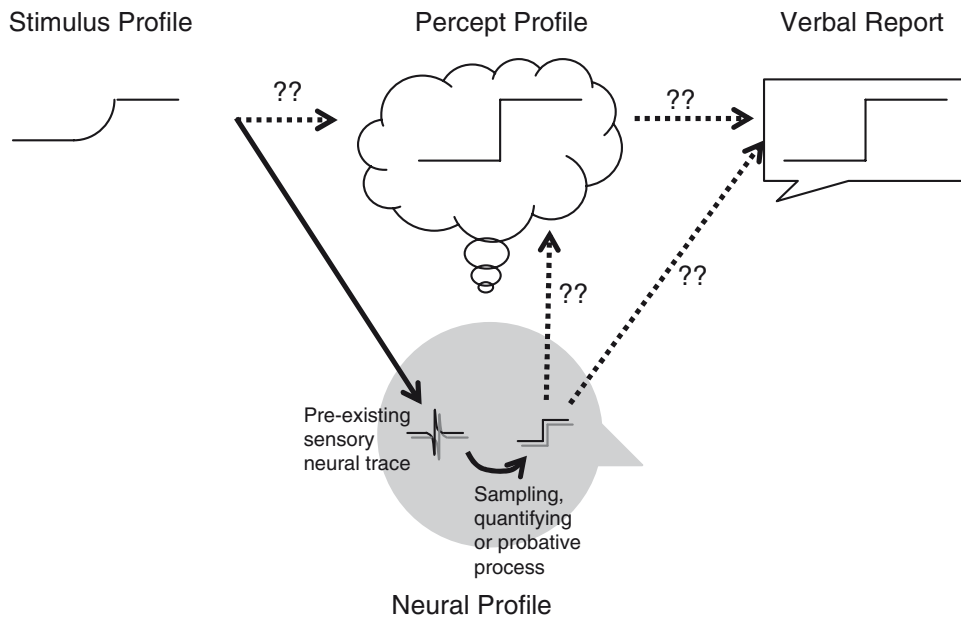
Fig. 8 *Complications Arising from Laming's Concerns*. If the report is a product of a process of a subject comparing stimuli, imposing a quantitative interpretation on a stimulus, or otherwise interrogating a pre-existing mental state, then one must suppose two distinct neural states, one a "raw" state that is present regardless of whether the subject performs this additional action, and one that is a result of this action. If the report faithfully reproduces features of any neural state, it is only this latter state, and not the "raw" state that is thus reproduced

by employing a similar representational system with different values, or else by employing a different representational system. This, however, presents us with a new difficulty. To the extent that the data we start with are the intensity of the stimulus and the report or reaction of the subject, it seems to me that we are now faced with a methodological problem. For all such an experiment gives us direct access to is the output of the sampling/judgment system. Its relation to the pre-existing "raw" state cannot be directly interrogated through this methodology.

One possible way of proceeding is to investigate whether giving subjects different types of instructions produces systematic differences in the psychophysical data produced. There is evidence that this is the case, including the mismatches between the results of different experimental methods in psychophysics. This is usually interpreted as a case of one method introducing experimental artifacts or errors that another avoids. But, more radically, it might be the case that *no* methodology of outer psychophysics can avoid the introduction of artifacts produced by sampling and judgment.

A second approach would be to move, as technology permits, away from relying on subjects' reports, and instead to compare stimuli with neural states – that is, to replace outer psychophysics with early psychophysics. This, however, has methodological difficulties of its own, of the sort brought up by Todorovic in his discussion

of isomorphist and non-isomorphist approaches. If we do not rely upon subjects' reports, how do we know when we have reached a stage at which the relevant questions have been answered, and that we are not stopping at too early a stage of sensory processing? On the other hand, if and when we do find a neural state whose profile matches that of the percept-as-reported, how do we know that what we are measuring is the "neural correlate" of the sensory state as opposed to that of the judgment about it, one that may misrepresent what it is a judgment about? This seems to leave us with a methodological dilemma. The data we possess for outer psychophysics are, if Laming is correct, obtained on the basis of subjects' judgments, and are not simply direct reproductions of pre-existing phenomenological or neural states that occur when the subject is *not* making a judgment. And these data play an important role in guiding our investigation of the neural processes underlying sensation and perception. But it is possible that the reports misrepresent properties of the phenomenological states themselves, and whatever neural states are specially related to *them*. If we discount the outer psychophysical data, we deprive ourselves of our principal way of mapping out the roles of different neural circuits in sensation and perception. But if we assume them to be faithful reports of the phenomenological states themselves, we ignore the very real possibility that judgment introduces an artifact into the reports upon which such data are based.

Philosophical Assessment of Laming's Challenge

Where does this leave us with respect to our prior assessment of the role of phenomenology in psychophysics?

First, it should be noted that, unlike some other critics of the role of phenomenology in psychophysics, Laming is not attacking a realist interpretation of sensory or perceptual states, understood from a phenomenological perspective. It is consistent with his arguments that there *are* subjective experiential states that are crucial to conscious sensation and perception. What he calls into question, however, are the assumptions (1) that the data acquired in psychophysics experiments amount to measurements of the subjective intensities of these states, and (2) that subjects' responses in such experiments are simply accurate reports of properties of pre-existing phenomenological states acquired through introspection. Laming argues that the first assumption is false; on the second he does not express a definite opinion, though we saw in the previous section that his work suggests some models on which the phenomenology of sensation and perception plays no causal role in the production of judgments about intensity. Let us pursue these issues in order.

Let us suppose that Laming is correct in holding that the data of psychophysical experiments collectively support only ordinal comparisons and not the imposition of the kind of ratio scale needed for the laws propounded by Fechner or Stevens. This is compatible with the possibility that subjects make their judgments on the basis of the subjective intensities of sensory *experiences*. However, it implies that if this is so, those subjective intensities may not be related to stimulus intensities by

Fechner's or Stevens's Laws. If the logarithmic and power-function interpretations of the data are consequences of artifacts introduced in the process of the subjects making quantitative judgments or the theorist concentrating on only a subset of the available data, psychophysical experiments do not provide us adequate evidence to reach any conclusion about the intensities of subjective qualities, or their relations to intensities of stimuli, beyond that of ordinal comparisons. We can perhaps conclude that *this* percept is more intense than *that* one, but not that it is *n* times as intense. The phenomenologist is free to endorse this, by holding that we are capable of ordinal comparisons of sensations *based upon their subjective intensities*, but that neither introspection nor the JND method provides the basis for imposition of a ratio metric.

The suggestion that judgments of stimulus intensities, or of the relative intensities of pairs of stimuli, play a significant role in the form the experimental data take, poses more of a problem for the phenomenologist. A great deal depends upon what is the right model of the causal pathways leading, on the one hand, to phenomenological states and, on the other, to the judgments that provide experimental data. If the judgments are based upon some type of introspection of phenomenological states, then phenomenology still plays a methodologically-crucial role in the collection of psychophysical data, even if the judgments introduce the illusion of being able to compare qualitative intensities algebraically. However, if the subjective intensities of sensations are not causally implicated in the production of such judgments, but are effects of a common cause, phenomenology plays no such role, and indeed subjects' reports provide only undependable and indirect evidence about the intensities of subjective states.

My assessment of alternative models of Laming's position also presents problems for the isomorphist methodological canon endorsed in earlier sections of this paper and in Horst (2005). If the process of judgment introduces artifacts into the data, then the data based on such reports may not accurately represent the subjective intensity contours of the phenomenological percept, *even if the latter plays a causal role in the production of the judgment*. If this is the case, then the psychophysical data may not in fact reflect properties of the percept, in which case the latter does not play the role I claimed for it in constraining theoretical models. Indeed, Laming would probably reject the isomorphist canon in any case, as he views some properties of subjects' responses as artifacts of the experimental task which result in misrepresentation of some of the properties of inner variables, whether phenomenological or neural.

How should we evaluate this threat? I think that we must do so on a case by case basis. Laming's arguments are largely concerned with general scales of sensation – the sort of thing that Fechner's and Stevens's Laws attempt to capture mathematically. It is less clear what implications they have for the kinds of things that concerned Todorovic, such as the claim that the percept involved in viewing the Cornsweet illusion involves a step-shaped rather than a cusp-shaped gradient of subjective intensities. Here, it seems to me, phenomenological methods *are* useful. We can, for example, examine the figure at greater length, asking ourselves whether brightness seems constant within a section of the figure, and whether the boundary

looks abrupt, ramped or cusp-like. This strikes me as (a) a methodologically-appropriate way of clarifying the data, (b) based in an examination of a subjective percept, and (c) of importance in constraining theoretical models that can accommodate the illusion. The subject is *not*, after all, simply making a careless error in her report. Rather, some features of the brightness contour of the stimulus have been lost at some stage of neural processing prior to the production of the percept. And this observation is a fruitful one: it tells us to look for a stage of visual processing that (a) has the contour of the percept, (b) is consciously accessible, and (c) is downstream of (or in a mutual feedback cycle with) neural units involved in the constitution of boundaries. And indeed, there are neural models that do just this. (Cf. Grossberg, *ops. cit.*)

Laming's model may also provide an additional way in which phenomenology is involved in psychophysics. Laming holds that the responses of subjects are based upon judgments. Judgments are intentional states. If, as some theorists (Searle 1992; Siewert 1998; Horst 1996; Horgan and Tienson 2002) have argued, intentional states have a phenomenology, and this phenomenology plays a causal role in subjects' reports, then the phenomenology of those judgments plays a crucial role in the production of psychophysical data, even if the process of judgment results in a misrepresentation of properties of the percepts themselves.

References

Boring EG (1921) The stimulus-error. Am J Psychol 32:449–471

Boynton RM (1989) About assumptions and exponents. Behav Brain Sci 12:271

Brentano F (1874) Psychologie vom empirischen Standpunkt: Duncker & Humbolt, Leipzig.

Campbell FW, Howell ER, Johnstone JR (1978) A comparison of threshold and suprathreshold appearance of gratings with components in the low and high spatial frequency range. J Physiol 284:193–201

Campbell FW, Robson, JG (1968) Application of Fourier analysis to the visibility of gratings. J Physiology 197: 551–566

Cartwright N (1983) How the laws of physics lie. Oxford University Press, Oxford

Chisholm R (1957) Intentional inexistence, *perceiving: a philosophical study*. Cornell University Press, Ithaca

Cornsweet TN (1970) Visual perception. Academic, New York

Davidson D (1970/1980) Mental events, *essays on actions and events*. Oxford University Press, Oxford, pp 207–224

Fechner GT (1860) Elemente der Psychophysik. Breitkopf and Härtel, Leipzig

Fechner GT (1877) In Sachen der Psychophysik. Breitkopf and Härtel, Leipzig

Grossberg S (1987) The adaptive brain. North-Holland, Amsterdam

Grossberg S (1994) 3-D vision and figure-ground separation by visual cortex. Percept Psychophys 55(1):48–120

Grossberg S, Mingolla E (1985b) Neural dynamics of form perception: boundary completion, illusory figures and neon color spreading. Psychol Rev 92:173–211

Grossberg S, Todorovic D (1988) Neural Dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception & Psychophysics* 43: 241–277

Horgan T, Tienson J (2002) The intentionality of phenomenology and the phenomenology of intentionality. In: Chalmers DJ (ed) Philosophy of mind: classical and contemporary readings. Oxford University Press, Oxford, pp 520–533

Horst S (1996) Symbols, Computation and Intentionality: A Critique of the Computational Theory of Mind. University of California Press, Berkeley

Horst S (2005a) Modeling, localization and the explanation of phenomenal properties. Synthese is 177(3):477–513

Horst S (2005b) Phenomenology and psychophysics. Phenomenol Cogn Sci 4:1–21

Horst S (2007) Beyond reduction: philosophy of mind and post-reductionist philosophy of science. Oxford University Press, New York

James W (1890) The principles of psychology. Henry Holt, New York

Laming D (1997) The measurement of sensation. Oxford University Press, New York

Lockhead GR (1992) Psychological scaling: judgments of attributes or objects? Behav Brain Sci 15:543–601

Ratliff F (1978) A discourse on edges. In: Armington JC, Krauskopf J, Wooten BR (eds) Visual psychophysics and physiology. Academic, New York

Ratliff F, Sirovich L (1978) Equivalence classes of visual stimuli. Vis Res 18:845–851

Savage CW (1970) The measurement of sensation. University of California Press, Berkeley

Searle JR (1992) The rediscovery of the mind. MIT, Cambridge, MA

Siewert C (1998) The significance of consciousness. Princeton University Press, Princeton

Stevens SS (1951) Handbook of experimental psychology. Wiley, New York

Stevens SS (1957) On the psychophysical law. Psychol Rev 64:153–181

Todorovic D (1987) The Craik-O'Brien – Cornsweet effect: New varieties and their theoretical implications. Perception and Psychophysics 42: 545–560

Tumarkin A (1981) A biologist looks at psycho-acoustics. Behav Brain Sci 4:207

von Kries J (1882) Ueber die Messung intensiver Grössen und über das sogenannte psychophysische Gesetz. Vierteljahrsschrift für Wissenschaftliche Philosophie 7:257–294

Weber EH (1834) Du pulsu, resorptione, audiu et tactu. Annotationes anatomicae et physiologicae. Koehler, Leipzig

Zuriff GE (1972) A behavioral interpretation of psychophysical scaling. Behaviorism 1:118–133

A Neurophenomenological Study of Epileptic Seizure Anticipation

Claire Petitmengin

Introduction

This article sets out to retrace the course of a neurophenomenological project initiated by Francisco Varela, the purpose of which is the anticipation of epileptic seizure, and to evaluate the relevance of the neurophenomenological approach from the methodological, therapeutic and epistemological viewpoints. New mathematical methods for analysing the neuro-electric activity of the brain have recently enabled researchers to detect subtle modifications of the cerebral activity a few minutes before the onset of an epileptic seizure. Do these neuro-electric changes correspond to modifications in the patients' subjective experience, and if that is the case, what are they? In a first part, after having recalled the context of the project, I will describe the methods I used for trying to detect the dynamic micro-structure of preictal experience, the difficulties I met and the results I obtained. Then I will show how the "pheno-dynamic" and neuro-dynamic analyses have guided, determined and mutually enriched each other throughout this project. In a third part, I will show that this genetic approach to epileptic seizure opens a new line of research into a cognitive and non-pharmacological therapy for epilepsy. Finally, I will argue through this example that neurophenomenological co-determination could shed new light on the difficult problem of the "gap" which supposedly separates subjective experience from neurophysiological activity.[1]

C. Petitmengin
Institut Télécom, Paris, France
Centre de Recherche en Épistémologie Appliquée (CREA), Paris, France
e-mail: Claire.Petitmengin@polytechnique.edu

[1]The structure and content of this article is partly inspired by Le Van Quyen and Petitmengin (2002), Petitmengin (2005), Petitmengin et al. (2006) and Petitmengin et al. (2007).

Context of the Project

The Neurophenomenological Program

The founding idea of neurophenomenology, a research program initiated by Francisco Varela (1996, 1997), is the following: in order to progress in the understanding of the human mind it is indispensable to try "to marry modern cognitive neuroscience and a disciplined approach to human experience, with respect to the continental tradition of phenomenology" (Varela 1996, p. 330). The neurophenomenological program is closely linked to the development of a *dynamic* approach to cerebral activity as well as to lived experience. On the one hand, the brain is considered as a complex system that cannot be understood by studying its different zones separately, but by highlighting its general (spatial and temporal) dynamics (Varela 1995; Thompson and Varela 2001; Freeman 2001; Le Van Quyen 2003). Detecting the dynamics involves the use of specific mathematical tools. On the other hand, cognitive events, far from being immediate and passive representations reflecting the properties of a pre-given external word, are considered as active processes that also develop in time. However, a dynamic, genetic description of experience also calls for special tools. Indeed a large part of our cognitive processing is pre-reflective,[2] that is unfolds below the threshold of consciousness. But it is possible to gain access to it thanks to very specific expertise that may be learned and perfected.[3] "There are numerous instances where we perceive phenomena pre-reflexively without being consciously aware of them, but where a 'gesture' or method of examination will clarify or even bring these pre-reflexive[4] phenomena to the fore." (Varela and Shear 1999b, p. 4). Neurophenomenology argues that unfolding this twofold dynamic is what allows the articulation and mutual enrichment of phenomenological and neurological analyses.

The Anticipation of Epileptic Seizures

Epilepsy is one of the most frequent neurological affections of children and adults: it affects about 1% of the world population. It is the consequence of a neuronal dysfunction that expresses itself by the sudden apparition of seizures.

---

[2]The term "pre-reflective" (in French "pré-réfléchi", to use the vocabulary of Husserl (1913), later adopted by Sartre [1936, 1938] and Ricœur [1949]) qualifies the part of our lived experience which, although "lived through" subjectively, is not immediately accessible to consciousness, introspection or verbal report.

[3]On the need to develop specific methods for studying lived experience: (Varela et al. 1993; Varela and Shear 1999a; Depraz et al. 2003; Petitmengin, 2009). For a synthesis of the difficulties of becoming aware of one's pre-reflective experience: (Petitmengin 2006b; Petitmengin and Bitbol, 2009).

[4]In this quotation the term "pre-reflexive" is equivalent to "pre-reflective".

An epileptic seizure is due to an abnormal and transient hyper-activity of cerebral activity. This hyper-activity starts in an area which is called the epileptic focus, and then expands to the neighboring areas and sometimes to the whole brain.

The unpredictability[5] of the seizures is the main reason for the poor quality of life of epileptic patients and leads to permanent insecurity for them and their families. In the past, this suddenness and the difficulty to identify the causes of seizures often led to the invocation of supernatural explanations. Today, in spite of important progress in the understanding of cellular mechanisms of epilepsy and in the localization of cerebral dysfunctions, we still do not understand why and when a seizure may emerge. Notably, until recently, no method for analyzing the neuro-electric signals could detect any changes before a seizure. The electroencephalogram (EEG), especially if it is recorded by means of intracranial electrodes which are in contact with the epileptic focus, locates very precisely the onset of a seizure. But before the critical discharge, no signal at the level of the raw EEG permits the anticipation of the emergence of a seizure. However, over the last 5–6 years, new "neuro-dynamic" methods for analysing the EEG enabled researchers to detect subtle modifications of the cerebral activity a few minutes before the onset of a seizure (2003; Mormann et al. 2000, Le Van Quyen et al. 2001).

Neuro-Dynamic Analysis of Seizure Anticipation

Let's describe these methods briefly. The neuro-dynamic analysis of cerebral activity relies on the following strong hypothesis: the emergence of a cognitive act does not correspond to the activation of a particular area of the brain, but to the activation of a multiplicity of spatially distributed regions, that coordinate their activities through a mechanism of integration. The implied mechanism is temporal resonance: groups of anatomically distant neurons communicate transitorily by synchronizing their oscillating activities in time (Varela et al. 2001). This neuronal dynamic is not organized in a sequential order, as the "computer" metaphor would have it. On the contrary, this hypothesis emphasizes the importance of the properties of networks having reciprocal connections, where the sequential character is replaced by a parallel process of network synchronization. One could take the analogy of an orchestra: suddenly, groups of distant instruments start playing on the same rhythm.

A network of synchronization is identified by analyzing the electroencephalographic activity of the subject: electrodes are placed at different points either on the surface of the skull, or inside the skull (intracranial electrodes directly implanted in the brain of epileptic patients for presurgical evaluation). Each captor measures the frequency of the different cerebral rhythms emitted by the population of neurons situated in this point of the brain. A synchrony for a given electrode pair occurs when two neural populations oscillate in a precise phase-relationship that remains

---

[5]In Greek, *epilambanein* means "to fall suddenly".

constant during a given number of oscillation cycles, in a given frequency range. This phenomenon is described as *phase-locking* (Lachaux et al. 1999). The phase synchrony analysis method consists of identifying the evolution of neuronal synchronization during the realization of a given cognitive process, in order to detect the possible corresponding succession of characteristic neuronal configurations (or "signatures").

Using this technique, Varela's research team performed an analysis of the synchronisation between intracranial recordings from eight patients with neocortical focal epilepsy being evaluated for epilepsy surgery (Le Van Quyen et al. 2001 and 2003). Synchronies between all possible combinations of electrode pairs in the hour before seizures for the entire frequency range (0–100 Hz) were examined.

This analysis identified three distinct phases (Fig. 1):

- The interictal[6] period, which is characterized by a lot of uniformly distributed synchronies.
- The preictal period: in 77% of the seizures, about 5 min before the seizure onset, a decrease in synchronisation of the neuronal populations surrounding the
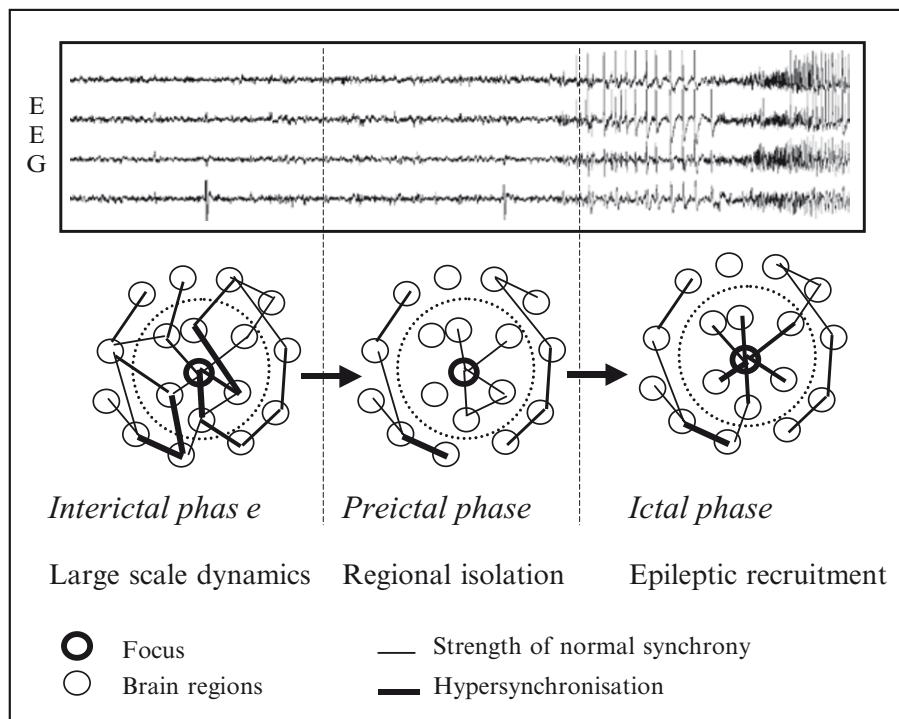
Fig. 1 The dynamics of neuronal synchronisations before and during a seizure[7]

---

[6] Inter-ictal (from Latin *ictus*: crisis) means "between two seizures". Pre-ictal means "before a seizure".

[7] This figure was created by Michel Le Van Quyen.

epileptic focus is observed within the 10–25 Hz range (the so-called beta 1 band). This decrease, which is the exact converse of phase-locking, is described as *phase-scattering*, where the probability of finding synchrony between two electrodes drops far below the interictal level (Varela et al. 2001). A decrease of synchronization in these populations with distant areas of the brain is also observed (leading to a "dynamic isolation" of the focus).

- On the contrary, the seizure is characterized by a higher synchronization of neurons surrounding the epileptic focus, while this area remains isolated from the rest of the brain.

These results show that the seizure does not arise suddenly, but that there is a transition from the interictal state to the ictal state. They also lead to the conclusion that the seizure does not correspond to the deficient functioning of a precise area of the brain, but to the deficient functioning of neural networks, related by abnormally facilitated connections. But the synchrony analysis does not tell us anything about the way this transition and this deficit are (or are not) felt by the patient. It indicates the *structure* of the cerebral activity, not the *nature* of the subjective experience that could correspond to it.

At the time of this discovery, I had just finished under the direction of Francisco Varela a doctoral thesis whose objective was to collect descriptions of the subjective experience that prepares and accompanies the emergence of an intuition. By means of specific interview techniques, I had collected the description of a variety of intuitive experiences. The analysis and comparison of these descriptions had enabled me to highlight a succession of very precise states and "inner gestures", showing a striking regularity from one experience to another and from one subject to another, in other words a g eneric dynamic structure of intuitive experience (Petitmengin-Peugeot 1999; Petitmengin 2001). Varela suggested that I then use the same type of approach to try to answer the following question: do the neuro-electric preictal modifications identified among epileptic patients correspond to modifications in their subjective experience, and if that is the case, what are they?

Pheno-Dynamic Analysis of Seizure Anticipation

Context

Concerning the patient's lived experience, the interictal and the preictal periods have not been much explored: the number of publications concerning them is small (Marchand and de Ajuriaguerra 1948; Gastaut 1954; Rajna et al. 1997; Schulze-Bohage et al. 2006) compared to the huge amount of work concerning manifestations related to the seizure itself, and to the so-called "aura". The aura (or simple partial seizure) corresponds to the sensations related to the irruption of a seizure, that is ictal manifestations, associated with visible electrical modifications on the

EEG. Auras are usually brief, lasting a few seconds or minutes (less than 5 min). Auras can be isolated, or can evolve into a complex partial seizure. As they provide crucial information in localizing epileptogenic zones where the seizure starts, neurologists and epileptologists are involved in identifying auras. On the other hand, preictal symptoms, also called prodromes, premonitory sensations, or warning symptoms, are less studied or acknowledged, probably because they seem to have no localization value. Moreover, their investigation has relied almost entirely on the use of questionnaires, which are not sufficient to obtain precise and reliable first person descriptions. These prodromes are the sensations I have investigated.

My objective is thus to help a set of epileptic patients to achieve a micro-temporal description of their subjective experience during the minutes preceding a seizure. And then to analyze these descriptions in order to try to detect the dynamic structure of this experience, and identify possible regularities. My goal is to explore the period that precedes – not the emergence of a moment of consciousness, as in the intuitive experience – but as is often the case in a seizure, a loss of consciousness.

For this project, the Neurodynamics Group from the Cognitive Neurosciences and Brain Imaging Laboratory[8] (Michel Le Van Quyen, Vincent Navarro, Jacques Martinerie and Francisco Varela), worked with a team of epileptologists from the Epilepsy Unit of the Pitié-Salpêtrière Hospital Paris (headed by Michel Baulac). This team put me in touch with epileptic patients. The selection criteria were (1) the ability to recognize a "high-risk" state for a complex partial seizure spontaneously, even without awareness of precise warning symptoms the description of which would be refined in the course of the interviews; (2) a capacity for introspection and self-expression which is sufficient to participate in a phenomenological interview. We excluded the patients whose elocution capacities were too much impaired. But we also noticed that belonging to a high socio-cultural category did not guarantee an easy access to one's lived experience, since the patient is more tempted to move from a description of the particular experience he has lived to a description of his representations and theoretical knowledge about this experience. Most of these patients suffer from a partial epilepsy of the temporal lobe.[9]

Within the space of 18 months I met nine adult patients, at their homes for two of them, during a stay in hospital for the others. The later were generally hospitalized for a pre-surgical exploration. They suffer from pharmaco-resistant epilepsy, for which no pharmacological treatment is efficient. In desperation, they resort to the ablation of the zone of the brain corresponding to the epileptogenic focus. In order to identify this zone as precisely as possible, through different examinations, a hospitalization of several weeks is necessary previous to the operation.

---

[8] Or LENA (CNRS UPR 640, Paris, France).

[9] A partial seizure (*versus* a generalized seizure) is a seizure which starts locally in the brain at the level of the epileptogenic focus and progressively recruits other cerebral areas, but does not expand to the whole brain. It is not always accompanied with a loss of consciousness.

The patients are hospitalized in a special room where they can be filmed and their intracranial EEG recorded 24 h a day. It is there that I met them.

Collecting Descriptions of the Preictal Period

Initializing the Interview

I noticed rapidly how important it is for me to take the time to introduce myself and to present my research. I proceed more or less as I have done in the present article: "My research group has detected neuro-electric signals before seizures, and my task is to verify if these signals correspond to a lived experience. Therefore I propose to ask you questions about the moments that precede a seizure. I would like to collect a precise as possible description of your experience during the period that precedes a seizure."

It is also very important, as in every phenomenological interview, but even more in this context, to create a relationship of confidence with the patient. I am going to require the patient to make a considerable effort to immerse himself back into the period having preceded a seizure, in other words a painful and harrowing episode. To agree to this, he must find me receptive, attentive, and compassionate. Before beginning the interview, I thus always spend a long moment inquiring about the circumstances of his hospitalization, and listening to what he is willing to confide in me about his illness. Once this relationship of trust is established, I try to encourage him to leave the doctor/patient framework, to leave for the time of the interview his position of "patient" to play the active role of co-researcher.

Another question soon arises: what is the patient's interest in participating in this research on seizure anticipation, what can motivate him? As soon as he understands the objective of this research, the patient starts imagining the hope of a new therapy, a possible recovery. But it would be cruel to give him the hope of a device enabling him to anticipate and even to stop his seizures: in the best of cases, this device will only be operational in several years. Moreover, most of the patients I meet have chosen the radical solution of surgery, and are relieved to have finally been called, after months of waiting, for the presurgical exploration. In these conditions, why participate in this research? For future patients? For science? The question remains unanswered, as I don't have any precise response to propose. And when I am facing a patient wearing intracranial electrodes, who is weakened, sometimes exhausted by the numerous seizures he has just experienced, I feel resourceless. A physician, a psychologist, comes to take care of him. But me, what can I bring to him?

Once the (immediate) objective of the interview defined, and the relationship established, the other elements of the contract are specified as usual for this type of interview: "If you agree, we are going to allow about 1 h to this interview, you should know that you do not have to answer all of my questions, and that you may also stop the interview at any time you wish, in this case just make a sign to me."

Main Difficulties

The main difficulty of the interview is the often pre-reflective character of the preictal sensations, which explains the paucity of initial verbal self-reports on preictal experience. As a large part of our lived experience, preictal sensations often do not reach the threshold of consciousness. Becoming aware of this pre-reflective part and describing it runs up against a set of obstacles. The most important of these obstacles are the dispersion of attention, its absorption into the content of the experience (the "what") to the detriment of its mode of appearance (the "how"), the lack of awareness of the dimensions and level of detail to be observed, and the necessity to access experience retrospectively. In addition to these obstacles which concern subjective experience in general, there are difficulties particular to preictal experience:

1. The belief, firmly anchored in the environment of the patients and reinforced by the medical discourse on epilepsy, in the suddenness and the unpredictability of seizures, which considerably hampers awareness of their early symptoms.
2. The fact that the perception of warning signals often triggers an emotional reaction of distress and panic, which in turn hampers the perception of these signals.
3. In addition, some patients may have permanent memory deficit or reduced ability to perceive subtle feelings preceding seizure, due to involvement of the medial temporal lobe in the epileptogenic zone, or due to the secondary effect of antiepileptic drugs.

In order to overcome these difficulties, I use a set of precise interview processes.[10]

Choosing a Singular Seizure and the Start of the Description

The choice of a seizure on which it is possible to work is an important and delicate moment of the interview. Because the seizures are often nocturnal: the patient is then unconscious during the preictal period. And even if the seizure takes place during the daytime, it often obliterates all memory of the preceding moments, and sometimes even the memory of having had a seizure (peri-ictal amnesia). Among the seizures the patient remembers, I thus ask him to choose a particular one.

If the patient sometimes feels warning sensations, it is a seizure in which these sensations were especially vivid; if he doesn't, it is a recent seizure, or an earlier one that he remembers particularly. Then we have to identify the right moment to begin the description. In the case of warning symptoms, we choose a temporal marker shortly before the start of these symptoms, and begin the description there. If the patient has not felt anything in particular, we concentrate on the period during which neuro-electric modifications have been detected, that is between 10 and 30 min before the seizure.

---

[10] For a general description of the difficulties of description and the interview techniques used to overcome each of them, see Petitmengin (2006b). For a complete description of one of these methods, the "explicitation interview" see Vermersch (1994/2003).

Evoking the Preictal Period

Because the seizures cannot be foreseen, the description of the preictal period through an interview can only be performed at a temporal distance. This difficulty is not encountered in all neurophenomenological projects, but it is particular to this one. For example, in the protocols consisting of asking a subject to complete a cognitive task, while recording his EEG, the description of the corresponding lived experience can be collected immediately after the realization of the task.[11] In another neurophenomenological study (in progress) of schizophrenic hallucinations, the interview can take place "in real time", the patient describing his experience as it is happening. But by definition the preictal period can only be identified as such after the seizure, which is unforeseeable: consequently the interview, in the best of the cases, can only be carried out several hours after the seizure.

It is therefore mandatory to guide the patient towards the "evocation", "refreshment" or "re-enactment" of the preictal period. In fact, whether the experience to be described is in the past or has just happened, this re-enactment is the key of the process of becoming aware of this experience (Petitmengin 2006b). In order to guide the patient towards this re-enactment, I help him to rediscover the spatio-temporal context of the experience (when, where, with whom?), and then with precision the visual, auditive, tactile and kinesthetic, olfactory and possibly gustatory sensations associated with the experience, until the past situation is "re-lived", to the point that it is more present than the interview situation. A set of precise objective clues enables me to verify that the patient is actually evoking the experience.[12]

At this stage, a difficulty particular to epileptic patients and linked to the evocation state arises: several patients are reluctant or even refuse to participate because they are afraid that the evocation of a past seizure might trigger a new one. This is the case in the following excerpt, where I nevertheless manage to obtain a description, but without any guarantee of reliability that the state of evocation usually provides:

N. Before the seizures, I am depressed, I feel bad.

C. I would like you to let yourself go back to a particular moment where you felt this sensation, here (at the hospital) or elsewhere.

---

[11] For example the protocol that involves presenting the subjects with a 3D illusion (Lutz 2002).

[12] These indicators are verbal, non-verbal and para-verbal. The verbal indicators are the use of "I", the present tense, the specific context indicators (place and time), the concrete and detailed character (as opposed to conceptual and general) of the vocabulary used. An example of a non-verbal indicator is the direction of the eyes: when the subject is reliving the past experience, he takes his eyes off the interviewer to look "into space", to the horizon. Concomitantly, the flow of speech slows, and the words are often cut with silences: these para-verbal clues are the sign that the subject is plunging into himself to make contact with the pre-reflective dimension of his experience. At the same time, metaphoric or deictic co-verbal gestures appear.

N. No, I refuse. I'm so afraid that I don't want to think of it. I have the impression that if I think of it, it's going to come back, it's going to trigger a seizure.

C. Without thinking of it too much, could you describe this sensation to me approximately? Does it occur in a precise part of your body?

N. Sometimes yes, it gets me at the top of the stomach, as if it was digging into me, under the ribs, it hurts me. It compresses me. I have difficulty breathing, I'm very hot (…)

Twice, the evocation of the instants having preceded a past seizure actually triggers a new seizure, with a patient suffering from "reflex epilepsy" (a particular type of epilepsy which is precipitated by specific external or internal stimuli). This case has been related in (Navarro et al. 2006).

Describing the Various Dimensions of the Preictal Experience

When the usual indicators show me that the evocation is sufficiently stabilised, I use appropriate questions to guide the patient's attention towards the various dimensions of his preictal experience. First, I help him describe the temporal unfolding of this experience, either from the marker which has been identified before the seizure – "What happens just after? What do you do then?" – until the onset of the seizure – "How do you know that the seizure is about to occur? How do you know that the seizure has started?" – or in reverse, from the start of the seizure. These questions, which flag the various moments of his experience without suggesting any content (Vermersch 1994/2003), enable the researcher to obtain a precise description without infiltrating his own presuppositions. This "content-empty" questioning enables the gathering of a first level of description, taking the form of a succession of sensations, internal states and operations.

In this sequence one might ask: which are the elements related to the emergence of the seizure, and therefore the ones to explore deeper? The difficulty is that I have no indication about the dimension of the subjective experience to which the detected neuro-dynamic modifications could correspond, therefore no track to explore. A few patients who are aware of warning sensations manage to guide me in the choice of the sequences to examine. Otherwise, it is often (and as soon as the first interview) the description of an action that enables me to guide the patient toward the awareness of a warning sensation:

P. At that point I get off my bike

C. Why do you get off your bike?

P. (Silence…) I know that I'm about to have a seizure

C. How do you know that you are about to have a seizure?

P. (Silence…) Because I feel this sensation of compression in my lungs…

C. How do you know that this sensation of compression announces a seizure?

P. Because I recognize it.

C. How do you recognize it?

P. Several times, I have felt this just before.

This is how I discover progressively, associated to the premonitory sensations, a set of (physical or mental) operations, the descriptions of which I also collect: very implicit tests enabling the patients to recognize the sensation as premonitory, actions achieved in order to protect themselves from the imminent seizure (getting off the bicycle) or even to try to stop it. The interviews and their analyses thus progressively provide me with some knowledge of the type of sequences which are interesting to explore, and towards which I will guide the patients' attention in the subsequent interviews.

Once this first level of diachronic description is collected, I encourage the patient to come back to the most interesting elements in order to go deeper into his description. This process may take different forms.

- When a sequence of sensations is detected in a given register, encouraging the patient to verify if this sequence is not accompanied by sensations from another register. For this purpose, I find it very useful to carry out, before the interview begins, a small training exercise to raise the interviewee's awareness of these different registers. For example, I encourage him to recall a memory of a holiday, and then to successively describe the visual, auditory, kinaesthetic, emotional, olfactory and gustatory dimensions of the memory. During the interview itself, this training will help the patient to adopt the "attention position" required for becoming conscious of the different sensorial registers of the preictal experience.
- Deepening the synchronic description of a particular sensation. For example, I will help the patient to turn his attention from the content of an image towards its structural synchronic characteristics, of which he usually has no reflective consciousness: the dimensions of the image, its localization in space, its constructed or remembered character, or the patient's egocentric or allocentric "perceptual position" inside or outside the scene. If it is a matter of describing a sound, I will draw the interviewee's attention to the generic characteristics of a sound: its volume, its tone, its distance, its direction and its persistence… If the interviewee talks to himself, as is often the case, is it with his own voice, or with another voice? From which direction does this voice come? A bodily sensation may in the same way be very precisely described in terms of intensity, location, or dimension.[13]
- Deepening the diachronic description of a sensation. For this purpose, I help the patient to turn his attention from a given sensation towards the dynamics of its appearance, its genesis: the different, generally very rapid phases that have preceded its stabilization; at every phase, the very rapid succession of interior micro-operations that he makes to recognize the sensation, evaluate it, and eventually stabilize, amplify, or discard it.

---

[13] The focusing questioning mode is very well suited for helping a person direct his attention to his bodily feelings, intensify perception of the feeling and describe it (Gendlin 1996).

For example, Christelle's preictal sensations were hampered by an intense sensation of panic, of nervousness all over her body. The interviews helped her to become progressively aware that this sensation of panic was just preceded by a "little dizziness", itself preceded by a sequence of subtle sensations that she finally summarized in the following way:

> There's this headache: a circle in front of me down to the cervix, and which presses in all around the head. Then a feeling of heat inside my body which rises from my stomach to my head. At the same time my heart accelerates. Then just after a little dizziness, like when one moves one's head quickly, but in this case without moving one's head. Right after there's this sensation of panic, of nervousness all over the body.

She also became aware of the sequence of subtle micro-actions and tests that she makes in order to try to interrupt an incipient seizure:

– I start breathing when I feel my heart accelerating. I breathe from the abdomen to try and calm my heart. I breathe in deeply and then I stop. If I feel that it helps I breathe normally again. If not, I breathe out quickly a little air and inhale deeply again.
– How do you know that it helps?
– I have the impression that it decompresses, it seems to make this area here (the chest) freer. The feeling of heat goes, there is more room. It's the sensation I get when I start to exhale which tells me if I should continue and then start breathing normally or again inhale quickly.

Throughout the interview, to detect pre-reflective sensations I rely on a set of precise non-verbal clues, such as eye movements and co-verbal gestures. Attentive observation of the eye movements[14] enables me to identify the sensorial register in which the patient is situated at a given moment, without necessarily being aware of this, and to draw his attention to this register. For example, if the patient looks upwards, it is probably because he is forming a mental image. An apposite question, such as "As you talk, you are looking up there (upwards and to the left). What are you doing inside yourself as you look in this direction?" will probably enable him to become aware of this image and describe it. Similarly, when the patient is looking down, this is often a clue that he is feeling a bodily sensation or emotion. An appropriate question will enable him to become aware of this.

The interviewer's prompts are also based on the observation of the gestures accompanying the words spoken (or substituted for the words spoken) in a pre-reflective way. For example, noticing a deictic gesture towards the chest may help him to draw the patient's attention to a warning symptom, with the help of a question such as: "What is happening for you in the middle of your chest?" It was the case for Christelle, who repeatedly passed her hand over her forehead, which I finally pointed out to her: she then became conscious of a sensation, which until

---

[14] Various papers have shown that eye movements precisely indicate the sensorial register used (Kinsbourne 1972; Galin and Ornstein 1974; Grinder et al. 1977; Ellickson 1983; Buckner et al. 1987).

then had been pre-reflective, of a "slight touch, like a breeze, a veil that lightly touches my forehead", which marked for her the onset of a seizure.

The average length of a phenomenological interview was one hour and a half. All patients were interviewed at least twice. The subsequent interviews enabled me to collect more precise and/or complementary descriptions, especially when seizures had occurred in the meantime. In fact, the process of becoming aware is an iterative one: the researcher retrospectively draws the patient's attention to his warning symptoms, enabling him to become reflectively more and more aware of them at the very moment when a seizure is about to occur, awareness which facilitates in turn the subsequent interviews.

In addition to these interviews, and in order to facilitate them, several strategies were used to increase the patients' awareness of their symptoms: (1) to use a logbook, in which patients may daily describe their inner mental states, and may report all preictal and ictal symptoms after each seizure; (2) to involve relatives in the description of the preictal period. We plan to use the following strategies in the future : (3) to show patients the video recordings of their preictal period, when they are hospitalized for long term EEG-video recordings; (4) to train patients in relaxation techniques in order to avoid the distress sometimes associated with preictal or ictal symptoms, which submerges the preictal sensations.

Analysing and Comparing the Descriptions

Once the descriptions are gathered, reorganisation and analysis are necessary to delineate and represent the structure of the experiences described. The main stages are the following:

- Reorganising the sequence of the description. The chronology of the process of awareness and that of the experience are not identical. When the patient re-enacts the experience for the first time, he provides quite a coarse "large mesh" description. He needs to go over it several times to become reflectively conscious of all the dimensions of his experience, and to provide a fine mesh description.
- Analyzing each description in order to extract the temporal unfolding of the experience, that is the precise sequence of sensations and possible mental or physical actions which constitute the experience (for example: "I feel this sensation of compression in my lungs/I recognize it/I get off my bicycle in order to protect myself").
- Comparing these sequences in several descriptions enables the researcher to detect possible regularities on different levels: synchronic, diachronic or functional. A regularity on the *synchronic* level is a sensation or action which is described several times with similar words. A regularity on the dynamic or *diachronic* level is a succession of similar sensations and/or actions. A *functional* regularity is a succession of sensations and/or actions which are different but realized with the same objective, for example for avoiding or stopping a seizure.

- The detection of such (synchronic, diachronic or functional) regularities enables the researcher to identify (synchronic, diachronic or functional) *generic structures*, which are progressively extracted from the initial descriptions through successive operations of abstraction.[15] For example, from several descriptions such as "I am looking at the scene through my own eyes, as if I was there", is extracted the generic synchronic structure "first person perceptual position". From several descriptions such as "When I woke up this particular morning, I felt fragile", is extracted the generic synchronic structure "negative internal state". From several descriptions such as "I feel this sensation of compression in my lungs/I recognize it/I get off my bicycle in order to protect myself" is extracted the following generic functional structure: "recognition of a warning sensation/ measure of protection". These are such generic structures, abstracted from the comparison of several descriptions, which are then correlated with neuronal signatures.

An Example of Neurophenomenological Circulation

Pheno-Dynamic Structure of Preictal Experience

This work of description, analysis and comparison led to the following results. All nine patients experience auras (ictal phenomena), six experienced prodromes (preictal phenomena). The auras were varied, depending on the suspected epileptogenic focus: vegetative ($n=7$), dysmnesic ($n=4$), psychic ($n=3$), sensory ($n=1$) or motor ($n = 1$). Where auras were usually described as "positive" since they often corresponded to motor, sensory or verbal hyper-activity, prodromic symptoms frequently corresponded to a decrease (of energy, of vitality), a lack (of concentration, of words, of physical balance) and were described as "negative". Whereas auras occurred a few seconds or minutes before the other ictal symptoms, the delay between prodromes and seizure was usually of several hours (until 24 h). Whereas auras were sudden and intermittent, prodromes were continuous and progressive.[16] The main prodrome, which was often described by the patients with very similar words, was an internal state of "tiredness", "weakness", "lack of energy", or "fragility" ($n=4$). Other patients described a feeling of distress ($n=3$), ill-being ($n=1$), or "loss" ($n = 1$). This state may be associated with difficulties in concentrating and speaking ($n = 1$), clumsiness ($n=2$), hypersensitivity to light ($n = 2$), noise ($n=1$) or other stimuli ($n = 1$) and with headache ($n=2$).

---

[15] For more details on these operations of abstraction (classification/instantiation, aggregation/ disaggregation and generalization/specialization) see (Petitmengin-Peugeot 1999).

[16] These findings are confirmed by other studies, for example (Rajna et al. 1997), (Schulze-Bohage et al. 2006).

Here are two extracts of interviews with patients who identified a prodromic state before most of their seizures:

> This can be 24 h in advance. It's in the whole body, I feel ill at ease, inside, it's constant, and it won't leave me until the fit has manifested. What I feel is… a little as if my body was abandoning me, therefore it isn't responding as quickly as usual, taking longer to carry out the orders I give it. I will get a pain in the head, it starts at the forehead, passes to the temples and goes as far as the back of the neck, like a circle around the head, and then down the neck. (…) Then it's speech. I have a tendency to slur words, stammer, lots of little things like that. But not all the time: I might be talking correctly then in suddenly, oops!, the words won't come, I'll start coughing, steady myself and it will pass. Other things also, like slightly feverish but all over the body. It's like I said a while ago, that my body won't respond as quickly as usual. And also a lack of energy. If someone suggested doing something, going out, anything, I'll say no, because I don't want to do anything. Because I don't feel well, because things aren't going well (…) I have no energy, no vitality, no punch.
>
> It's difficult to explain. The course of my thoughts, at a given moment, concentrates suddenly all on... how to say that... on... it is a kind of global situation of what we are, in what we are, in what we go to. And it's rather negative, because all of a sudden, I realize the absurdity of what we are, of what we do. (...) I feel a sort of acceleration of my thoughts. There are many more things that turn more quickly in my head. I go round much more ideas at the same time, as if they are colliding, and if the set of things that collide can't find any resolution. And it's that that projects me into this vision of the absurd. It is the sign that I am very likely to have a seizure, that I am on the verge of a perimeter of seizure.

It's important to note that this sensation of fragility and loss, usually perceived as painful and harrowing, may also be experienced as a relief:

> When I began to have seizures, I felt a sort of pleasure falling into the seizure. The pleasure of someone who has dropped everything, his work, his troubles, his domestic problems… and who is finally going to recover the freedom of everything.

In the following description, Dostoievsky (who was epileptic) ascribes to his character the prince Mychkine an even more intense feeling of deep understanding, of quietening, of fullness:

> He was thinking, incidentally, that there was a moment or two in his epileptic condition almost before the fit itself, when suddenly amid the sadness, spiritual darkness and depression, his brain seemed to catch fire…his sensation of being alive and his awareness increased tenfold…his mind and heart were flooded by a dazzling light…culminating in a great calm, full of serene and harmonious joy and hope, full of understanding and the knowledge of the final cause.[17]

Therefore the preictal state seems to correspond to a very deep modification of the patient's relationship to himself and to the world, usually something experienced as very harrowing, more rarely as a liberation. This state lasts, often intensifying, until the onset of the seizure.

---

[17] Fyodor Dostoyevsky, "The Idiot" (1868).

Countermeasures

Seven of the nine patients, once the warning symptoms are detected, spontaneously adopt different types of cognitive or behavioral countermeasures in order to avoid or interrupt the seizure. The countermeasures are of a variable nature. Physical countermeasures consist of motor (getting up, walking), sensory (rubbing the area where the sensation occurs), or vegetative (regular breathing) activities. Mental countermeasures consist of resting and relaxing, concentrating (on an object, a music, or one's own internal monologue), talking, or imagining[18] (e.g., a pleasant situation). These countermeasures are initiated by the patient, or by relatives when they detect a seizure. As the goal of a phenomenological interview, unlike that of a questionnaire, is not to collect quantitative estimations, but very precise descriptions of *a few* preictal experiences, we cannot give a precise evaluation of the success of these countermeasures. Nevertheless, we gathered some testimonies which reflect the beliefs of the patients about the effectiveness of their countermeasures, such as: "When a seizure is arriving, my friend tells me funny stories. For little seizures, it always works." Or "Sometimes, I can stop the seizure with this abdominal breathing, that I learnt when I was practicing yoga. But most of the time, it only delays the onset of the seizure." The earlier the awareness of the warning symptom is, the more efficient the countermeasure seems to be.

In conclusion, this pheno-dynamic analysis shows that seizures do not always occur like a bolt from the blue: they are the result of an (often pre-reflective) microgenesis.

Articulating Pheno-Dynamic and Neuro-Dynamic Structures

Let's remember the initial question: do the neuro-electric preictal modifications identified among epileptic patients correspond to modifications in their subjective experience, and if that is the case, what are they? Our initial approach consisted in trying to identify a generic structure of the preictal subjective experience, and then to correlate it with the corresponding generic neuro-dynamic structure (Petitmengin et al. 2006).

Search for a Temporal Coincidence

First, we showed a time-lag: the decrease of neuronal synchronization occurs a few minutes before the seizure, whereas the state of fragility that seems to characterize the preictal period for the patients that we interviewed, is felt several hours before

---

[18] The epileptologist R. Efron (1956; 1957) explains that one of his patients managed to stop her incipient seizures by smelling a certain perfume. As her profession of singer did not allow her to have her bottle of perfume with her all the time, she learned to associate the smell of the perfume to a visual stimulus (a bracelet) first present and then only imagined, with the same result.

the seizure. The detection of this state suggested that the electro-encephalogram should be analyzed in the long term. An analysis of the evolution of cerebral synchronization has thus been achieved from the intracranial recordings of a group of five patients suffering from epilepsies of the medial side of the temporal lobe, and having been monitored continuous EEG-video recording 24 h on 24 (about 50 seizures and 305 h of recording were studied). This analysis enabled the identification of a "preictal state" characterized by a desynchronisation of the neuronal populations related to the epileptogenic focus, up to 5 h before the seizure onset (Le Van Quyen et al. 2005). These observations have been confirmed by other studies (Mormann et al. 2003a, b).

The neuro-dynamical analysis, here guided by the phenomenological analysis, seems to show that the seizure is just the "tip of the iceberg", the climax of a process that starts long before.

Search for a Structural Correspondence and Working Hypothesis

Prodromes were frequently referred to as "negative" symptoms, which subsist while intensifying until the onset of the seizure, whereas auras were frequently referred to as "positive" symptoms, appearing brutally and lasting a short time. We hypothesized that the subjectively negative character of the preictal state and the progressive increase in its intensity, observed among five patients in this study, could be correlated with the progressive loss of phase synchrony observed during the preictal period, and therefore be the clinical expression of the phenomenon. On the other hand, the "positive" and sudden character of the symptoms related to the aura seemed to correspond to the hyper-synchronisation observed as soon as the onset of the seizure. This concomitant evolution led us to hypothesize a homeomorphism between the preictal neuro-dynamic and pheno-dynamic structures (see Fig. 2).

Consequences and Lines of Research

> Exploring the pre-reflexive represents a rich and largely unexplored source of information and data with dramatic consequences (Varela and Shear 1999b, p. 4)

Therapeutic Consequence: A Cognitive Therapy of Epilepsy

Becoming aware of the pre-reflective genesis of a seizure opens a temporal interval during which the patient can take some countermeasures to avoid the seizure, to stop it, or at least to protect himself from it. This leads to an unexpected research line towards a non-pharmacological, cognitive therapy of epilepsy, and maybe another understanding of this disease.

The few research groups which have organized such programs of seizure control show promising results (Dahl 1992; Wolf 1997), since the reduction of the number of seizures is equal or even superior to the reduction obtained with the more efficient pharmacological treatments. For example, one of these programs has been carried out with 16 children: the treatment lasted between 3 and 30 months according to the children. Sixty eight percent of the children had a reduction of 80–100% of the number of their seizures, twelve percent a reduction of 60–70% (Schmid-Schönbein 1998). In the program described by Reiter and Andrews (2000), 79% of the 45 patients succeeded in controlling their seizures completely.

The setting up of countermeasures, like the detection of preictal symptoms, is not immediate, it requires long-term training. Moreover, epilepsy often seems to fulfill a function in the life of the patient and his whole family (e.g., see Soulayrol 1999; Diebold 1999). The cognitive treatment must then be accompanied with a deep reorganization of his life and his familial and social relations. But this treatment gives him the possibility to take control of his life, until now under the sword of Damocles of the seizure. The patient learns to establish a relationship with his seizures in a different manner: to observe them instead of only dreading them, to familiarize himself with them; and in a way to become reconciled with them. He learns to become an actor in his relationships with others, starting with his physician: indeed he plays an extremely active role in the specification of the treatment, and at the end of the treatment, he will control his seizures alone. This therapy initiates for the patient a deep transformation that is a key to his recovery.[19]

What is the role of the neurophenomenological approach in designing these therapies? This approach, rather than resulting in a reduction of the disease to an uncontrollable cerebral dysfunction, contributes on the contrary to highlight the possibility of controlling this functioning. First, because the identification of a particular neuronal configuration before the onset of a seizure reinforces the hypothesis of the existence of preictal sensations, which until now were not taken seriously and little studied, and therefore gives a neurological confirmation to the possibility of cognitive therapies.[20] Secondly, because the neurophenomenological approach helps to enrich and refine these therapies:

– Through the development of suitable interview techniques, which enable the therapist to guide the patient towards becoming aware of his preictal symptoms and describing them, and then specifying countermeasures
– Through the setting up (in progress) of biofeedback devices based on cerebral synchrony analysis, which facilitates the awareness of premonitory sensations and the setting up of countermeasures

---

[19] On this topic the reader may refer to the very interesting testimony of J. Benak (2001).

[20] There are other contexts where a neurophenomenological correlation provides a neurological, "third person" confirmation of the existence of sensations or states that are not taken seriously and little studied. It is for example the case of some deep meditative states. Recent research carried out on very experienced Buddhist meditators shows that these states have quite characteristic neuronal signatures, which have never been observed in a non-pathological context (Lutz et al. 2004).

Finally, the neurophenomenological articulation provides the beginning of an explanation for the efficiency of countermeasures, and a clue for choosing them. Results of phase synchrony analysis lead to the hypothesis that countermeasures are effective because they prevent the isolation of neurons in the epileptogenic focus using recruitment of surrounding neurons or a more global re-synchronisation of distant areas of the brain. The countermeasures must be chosen accordingly: for example, if the focus is related to the centre of language, the patient will be incited to talk. Further testing of this hypothesis would require a systematic search for the neuronal correlates of therapeutic countermeasures, a future line of research for the neurophenomenological analysis of seizure anticipation.

Looking beyond epilepsy, becoming aware of the pre-reflective micro-genesis of a given cognitive event develops the possibility of transforming this process. To perceive, memorize, imagine or observe, we undergo a sequence of very precise operations, usually completely pre-reflectively: subtle transformations of the direction, intensity, scope and source of attention, modifications of the perception position, appreciation, comparison. Only the later phases of this process usually appear to be conscious. Becoming aware of the primitive phases, of the microgenesis of this process, is like introducing a kind of 'game' or space that enables us to transform it. For example, developing an early awareness of the subtle sensations that announce the emergence of an emotion, and of the interior micro-gestures that amplify it and maintain it, enables us to learn to pacify and to foil the emotional process before the intensity of the emotion has become a possible cause of suffering. Or alternatively, I am not condemned to have a "poor memory": once I have become aware of this, I can transform the very precise sequence of inner micro-operations that I carry out to memorize or to remember. The anticipation of epileptic seizure seems to constitute a paradigmatic example of the possibilities of transformation that becoming aware of the pre-reflective micro-dynamics of lived experience develops.

Epistemological Implications: The Question of the "Gap"

As soon as its first formulation by Francisco Varela (1996), the neurophenomenological program claimed to shed new light on the explanatory gap[21] which separates objective biophysical processes from lived experience.[22] What sort of light do our

---

[21] See Levine (1983). The "explanatory gap" is the core of the "hard problem" formulated by David Chalmers (1996): *why* does conscious experience emerge from some neurophysiological processes. The reader will find a presentation of this problem in (Roy et al. 2002). The recent article of M. Bitbol (2006) contains a summary of the theories – Behaviorism, Identity theory, Eliminativism, Functionalism, Idealism – that try to solve the problem, without providing any satisfactory solution.

[22] "We need to turn to the only link between brain and consciousness that seems both obvious and natural: the structure of human experience itself" (Varela 1996, p. 330).
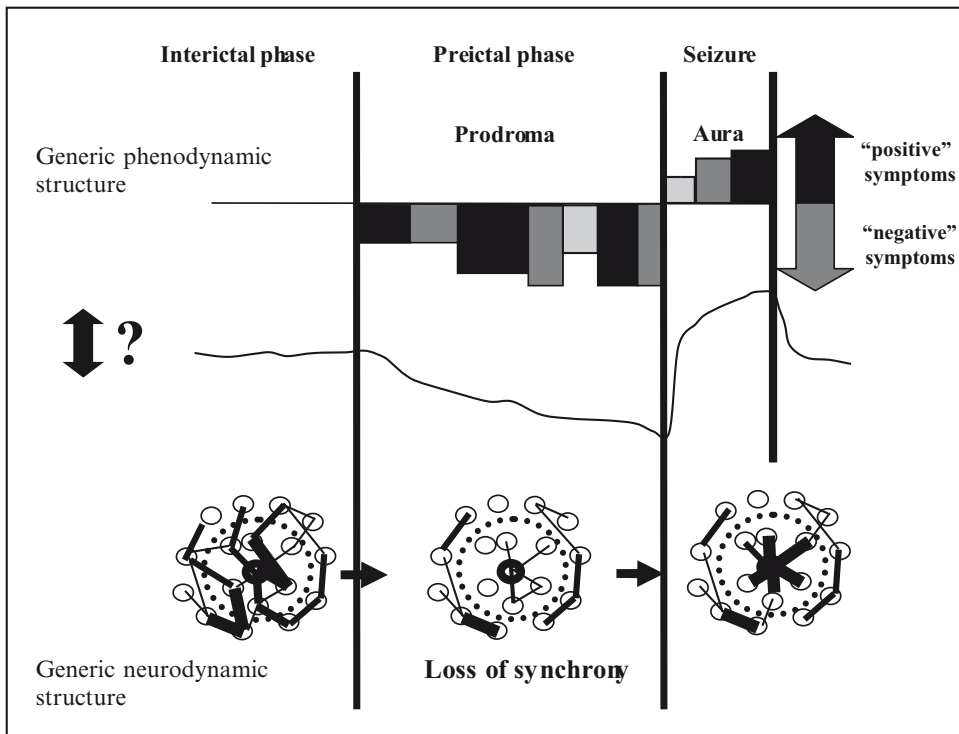
Fig. 2 Articulating pheno-dynamic and neuro-dynamic structures[23]

attempts to articulate pheno-dynamic and neuro-dynamic preictal structures shed on this question? (Fig. 2)

Searching for Homeomorphisms

Comparing phenomenal and neuronal structures enabled us to hypothesize a homeomorphism. This homeomorphism between the progressive decrease of preictal synchronization and the intensification of the "negative" state felt before the seizure, permits us to suppose a link between the particular quality of this experience – loss of consistency, of energy, of meaning – and the observed desynchronisation. *Through the intermediary of a homeomorphism of their dynamic structures*, which are themselves the products of a succession of complex transformations (i.e. the progressive extraction of generic structures), we may have established a footbridge between cerebral activity and subjective experience.

But while achieving these transformations, we have not reduced one level to the other. The quality of lived experience is not included in its dynamic structure:

---

[23]This figure was drawn by Vincent Navarro (Petitmengin et al. 2006).

for example, the specific quality of "the feeling of fragility I felt when I woke up yesterday morning, that has been increasing until the onset of the seizure last night", is not included in the generic dynamic structure: "intensifying negative internal state". In fact, the quality of lived experience is lost as soon as this experience is described: it is not sufficient to read or to hear the description of an experience to access the corresponding experience (in our example, to know "what it is like" (Nagel 1970) to feel the imminence of an epileptic seizure). From an experience to its description, a dimension is irremediably lost. At the very most, a description can trigger the described experience, or enable me to recognize it, if I have already lived it. In the same way, cerebral activity cannot be reduced to its dynamic structure, nor to its neuro-electric activity. Moreover, cerebral activity itself is only one element of a much more complex system, of which some dimensions are probably still unsuspected, implying not only our body but our whole environment. In other words, "the mind is not in the head" (Varela 1999a; Thompson and Varela 2001). Therefore the footbridges that we are establishing between the neurobiological and the phenomenal sides of the gap only enable us to begin to detect from one side the echo of the rhythms of the other. They enable us to anticipate the seizures better, and to understand[24] the preictal dynamics better. But they still don't enable us to explain how the peculiar quality of "this sensation of the absurd, with a very characteristic texture, that marks for me the perimeter of the seizures" can emerge from a neuronal desynchronisation. Bringing the pheno-dynamic and neuro-dynamic structures closer does not permit us to eliminate the distance that separates the structure of experience from its nature.[25]

If in the future the designing of suitable protocols enables us to identify the neurological correlate of the cognitive countermeasures for controlling seizures, and the dynamical structure of the process of interruption of an emergent seizure, correlating the neuro-dynamic and pheno-dynamic structures would enable us to build another type of footbridge. We could hypothesize, through the intermediary of the corresponding pheno-dynamic and neuro-dynamic structures, that a conscious cognitive act is one element in *constraining* cerebral activity. Thompson and Varela (2001) called this constraint *global-to-local determination* or *downward causation*. Such investigations are interesting to develop since most research in cognitive sciences relies upon the presupposition, and seeks to prove, that cerebral activity determines subjective experience, but not the inverse. However, highlighting this link would still not permit us to close the gap, that is to explain the phenomenal character of a cognitive act by the corresponding neuro-dynamic structure.

---

[24] Indeed, a homeomorphism between A and B brings elements of *explanation*, in the sense that "A is linked to B by a law". This doesn't mean at all that A is reducible to B, or B to A. And this link is symmetrical (if A is linked to B, B is linked to A).

[25] On this topic we fully agree with Baynes (2004, p. 13): "Formal models can only capture the structure of a domain; they cannot capture its intrinsic nature. Those who think that the hard problem is hard do so because they think that the phenomenal character – the 'what it is like' of experience – cannot be fully captured by structural descriptions".

The rough homeomorphism that we have identified could nevertheless be made more precise, thanks to finer comparisons of the neuro-dynamic and pheno-dynamic structures. This would imply several prerequisites:

– Finding a unique, intermediate and neutral formalism (symbolic or analogical), that would permit the representation of both neuro-dynamic and pheno-dynamic structures (this being especially difficult for the latter) (Varela 1997).
– Identifying more precisely the dimension of subjective experience which could conceal the echo of the cerebral synchronies.

Indeed synchrony analysis permits the discovery of some very subtle variations of rhythm, and modifications of the cerebral synchronization at very short intervals (a few hundreds of milliseconds). Which dimension of subjective experience could be correlated with such subtle and rapid variations? An investigation of the following ideas could help answer this last question.

The Rythmic and Transmodal Dimension of Lived Experience

Exploring the micro-temporality of lived experience, thanks to the refinement of first and second person methods, permits the discovery of deeply pre-reflective dimensions of this experience: fine sensations, subtle variations of internal rhythms, that correspond to the early stages of processes of which only the later stages usually appear to consciousness. It is for example the premises of an emotion, before it becomes intense and almost "solid", or the presentiment, the "interior direction" that announces the emergence of an idea, or a solution. These subtle "felt meanings" (Gendlin 1962) seem to unfold in an unrecognized dimension of our experience, whose structure is quite particular, very different from its more superficial structure.

First, the frontier between the different sensory modes is far more permeable than in our more conscious experience. Indeed the submodalities which character-ise this dimension – movement, intensity and rhythm – are "transmodal", that is they are not specific to a particular sense, but can be transposed from one sense to another[26] (unlike for example temperature and texture which are specific to touch, or volume which is specific to hearing). These subtle rhythmical modifications which Stern (1985) calls "vitality affects", constitute the internal world of the infant, and seem to remain active throughout adult life: they are the very texture of our experience.

This transmodality is associated with a transformation of the frontier usually perceived between the interior world and the outside world, which is described as far less rigid or even absent. The idea, the sensation, the memory emerges in a space

---

[26] Plato and Aristotle had already identified these characteristics, which they called "common sensibles".

which is neither subjective nor objective. In a concomitant manner, the feeling of individual identity changes: it becomes "lighter" or even disappears. For example, the descriptions that we collected of the instants of emergence of an "idea" – the solution of a problem, a new scientific idea, psychotherapeutic insight – show that the "sense of agency", that is "the sense that I am the one who is generating a certain idea in my stream of consciousness", as well as the "sense of ownership", that is "the sense that this idea is *my* idea" (Gallagher 2000) are often deeply altered (Petitmengin-Peugeot 1999; Petitmengin 2001).

We may find the same type of structure in the experience which accompanies the unexpected emergence of a memory. For a few instants, we are overwhelmed by a feeling which does not belong to a specific sensorial register, but is nevertheless intense and full of meaning, quite specific, and accompanied by a feeling of loosening of the limits of "self".

Even the emergence of a perception seems characterised by an initial instant, very rapid and usually completely pre-reflective, where the internal world and external world, the subject and object, are still indistinct. This instant is easier to recognise for a tactile or auditory perception. It is easier to recognize when you are surprised, or when you are awakening, or when you are very relaxed, for example while walking in the forest. A sound occurs, and for an instant, you do not know who you are, where you are, you do not even know that it is a *sound*. It's only an instant of consciousness hanging in the air, which may nevertheless be very intense and clear.

Whether it is a perception, an idea or a memory, these initial instants of lack of differentiation seem to be immediately followed by tiny gestures of distinction, separation, and then a very rapid succession of movements of identification, recognition, localisation, appreciation. In the course of this process, the emergence of the object and the emergence of the "self" seem concomitant. As Varela explains: "The borders between me and the others, even in the case of perception, are not clearly traced, and being a 'moi' and becoming a 'toi' are concomitant events." (Varela 1999b, p. 15). The more solid and stable the object becomes, the more "my" existence confirms itself. This mutual confirmation, which originates in tiny initial gestures, continues, at coarser levels, by means of discursive,[27] conceptual and emotional devices which are more easily accessible to consciousness.

It is in the microdynamisms animating this primitive dimension that moment after moment our feeling of identity and our relationship to the world seem to be played out. The epileptic seizure, often described as the dramatic loss of all points of reference, the downfall of the world and the loss of identity, could originate in this very deep dimension of our experience, from a disruption of this original rhythmic process.

---

[27]As the unceasing internal dialog, recognized as essential in the constitution of the "narrative self" (Gallagher 2000)

More generally, it is in this felt dimension, close to our body but made of a more subtle texture, which seems to be situated at the hinge between the physical and the psychic, that we could discover the subtle rhythms which could be correlated to the cerebral rhythms. We try to collect a very precise description of this dimension, which until now has been little explored because it is deeply pre-reflective and thus particularly difficult to become aware of, and to identify its dynamic microstructure (Petitmengin 2007).

Taking this dimension into account would probably enable us to establish more precise homeomorphisms between the neuro-dynamic and pheno-dynamic structures, and therefore to bring closer the two sides of the gap. But we don't think that bringing closer two structures, even if this is achieved at a very fine level of temporal granularity, permits the elimination of the distance that separates the structure of a domain from its intrinsic nature or "what it is like". The pheno-dynamic quality of lived experience is not fully contained in a formal structure, precise though it is. A structural homeomorphism permits us to confirm, to foresee, to understand, but not to eliminate this gap.

Questions of Co-Constitution

Another track in exploring the "gap" (complementary to the first) opens up at another level. It does not consist in trying to bring the neurological and phenomenological structures closer, but in observing, at a meta-level, the way these structures construct themselves, that is the *neurophenomenological circulation process* itself. In other words, it does not consist of detecting and comparing specific preictal states at the neurological and phenomenological levels, but of observing this very process of detection and comparison. It is not a matter of analyzing the epileptogenesis, but the genesis of this analysis itself.

First, this meta-observation shows that the neurological analysis and the phenomenological analysis guide and enrich each another. Indeed, the discovery of a new neuro-dynamic structure (the preictal neuro-electric desynchronisation) permitted a refinement of the awareness of the corresponding experiential dynamics (preictal symptoms and therapeutic countermeasures). Reciprocally, a refined awareness of the experiential dynamics enabled the detection of an original structure in the neuronal dynamics (neuronal desynchronisation at a distance from the seizures): in this second case, it is the phenomenological analysis that guided the neuro-dynamic analysis. In a different research context, these results confirm the heuristic character of the research strategy inaugurated by Antoine Lutz,[28] which consists of discovering an unobserved structure in the neuronal data by using phenomenological categories (Lutz et al. 2002; Lutz 2002; Lutz and Thompson

---

[28]This neurophenomenological research of Lutz deals with 3D vision.

2003). We notice that in both cases, it is the temporal unfolding of the phenomenological description (becoming aware of the *genesis* – of a 3D perception, or of an epileptic seizure) that enables the refinement of the neuro-dynamic analysis.

The neurophenomenological method therefore doesn't consist here of a simple "static" comparison of results that would be achieved independently. But it doesn't limit itself either to a heuristic process, the discovery of a regularity on one side triggering investigation on the other. We observe a real interweaving of the construction processes (that Francisco Varela (1997) called "mutual generative constraints"): the results achieved on one side not only trigger, but *structure* the process of analysis and the results achieved on the other side. This intertwining of the pheno-dynamic and neuro-dynamic analyses is especially well highlighted in the protocol of Lutz, where it is the use of an experiential category as a criterion for neuro-dynamic analysis that enables the detection of an original structure on this level (which confirms in turn the relevance of that category). Far from being constructed independently, the phenomenological and neuro-dynamic structures result from a complex process of mutual stabilization, selection, adjustment and validation, from a real "dynamics of reciprocal elaboration of the phenomenological and physico-physiological sides" (Bitbol 2006).

For example, a dynamic refinement of neuronal analysis contributes to making the patients conscious of the pre-reflective premises of their seizures. Therefore the effect of this analysis is to transform not only our *knowledge* of the preictal experience (which from nothing, gains the status of object of personal and scientific investigation), but this experience itself. The neurological analysis contributes to enriching not only the description, but the lived experience of the patient himself: it gives him access to a pre-reflective part of his experience that until now was inaccessible to him, and while enabling him to control his seizures, this consciousness will deeply transform his existence.

Conversely, the dynamic refinement of phenomenological analysis permits the discovery of a succession of characteristic neuronal configurations or "signatures", where until now only noise was perceived. Are these structures "discovered", or "built"? Can one consider that these structures existed in the brain before the researcher detected them? On the contrary it seems to us that neuro-electric activity is not "given", but that it is the result of a very complex construction process, of which every stage is determined (in a pre-reflective and implicit manner) by the degree of consciousness that we have of our subjective experience,[29] and limited and constrained by the previous stages of the process.

---

[29] Is there even one experimentation that makes absolutely no reference to subjective experience? For example, a cerebral area is called "auditory" because a correlation is made between its activation and a first person report of an auditory experience. The first person point of view is always present, but in an implicit, naive manner.

Towards the Origins of the Gap

Do these two neurophenomenological approaches: exploring a hinge dimension where the subjective and objective rhythms come closer, and observing the co-determination of the phenomenological and neurological sides, have a point of articulation? We hypothesize that these two approaches are ways of considering an identical process, the process of co-constitution of the subjective and objective poles, considered at two different levels:

1. The primitive and fine level of the microgenesis of a cognitive event: how, from an initial instant of lack of differentiation between the sensorial modes and between inside and outside, the "fission of Being" (Merleau-Ponty 1964, p. 81) is created and maintained; how do the subject and the object co-emerge, differentiating themselves while constituting each other, moment after moment, by means of minute gestures of distinction, separation, identification, recognition, localization, appreciation (Petitmengin 2006a).
2. The later and more elaborate level of the co-constitution of the phenomenological and neuronal sides (Bitbol 2006).

The question of the gap is thus being transformed: it is no longer enough to explain the lived experience by its hypothetical neuro-physiological substratum, but to understand the process of co-determination of the objective and subjective poles, a process of which neurophenomenological circulation is only a late instance. The question is no longer exhausting ourselves bringing closer the two sides of the gap – reducing one to the other, eliminating one for the other, or explaining one by the other, but to adopt quite a different point of view: to observe how the gap constitutes itself, and the different stages of this constitution. In this perspective, the question is not to try to eliminate all trace of subjectivity, but to observe how lived experience intervenes at the different stages of the co-construction of the two sides. The neuro-phenomenological correlations do not aim at trying to suppress or to overshadow the gap, but at revealing the synergy of the two sides, specifying and amplifying it.

Conclusion

The example of epileptic seizure anticipation shows that far from reducing one pole to the other, the neurophenomenological approach favors and amplifies the mutual refinement and unfolding of the subjective and objective dimensions. This double unfolding has two main consequences. On the one hand, by highlighting unsuspected dimensions of lived experience, it leads to a possible transformation of experience. On the other hand, it highlights the usually implicit process of co-constitution of the subjective and objective sides of the gap. It draws our attention to this dynamics and invites us to observe all its phases, from the later and coarser phases of neurophenomenological co-determination to the earlier and more subtle phases of the emergence of a moment of consciousness: how, instant after instant,

is the scission between the "self" and the "world" created and maintained? Let's imagine that a refined awareness of this process of co-constitution could profoundly renew our vision of ourselves and of our relationship to the world.

References

Baynes T (2004) Closing the gap? Some questions to neurophenomenology. Phenomenol Cogn Sci 3(4):349–364

Benak J (2001) To quantify experience? Methodological issues in the behavioural/psychological treatment of complex partial seizures. Seizure 10:48–55

Bitbol M (2006) Une science de la conscience équitable. L'actualité de la neurophénoménologie de Francisco Varela. Intellectica 1(43):135–157

Buckner W, Reese E, Reese R (1987) Eye movement as an indicator of sensory components in thought. Journal of Counseling Psychology 34(3):283–287

Chalmers D (1996) The conscious mind: in search for a fundamental theory. Oxford University Press, New York

Dahl J (1992) Epilepsy. A behavior medicine approach to assessment and treatment in children. Hogrefe & Huber, Seattle/Toronto

Depraz N, Varela F, Vermersch P (2003) On becoming aware. John Benjamin Press, Amsterdam

Diebold G (1999) L'épilepsie, une maladie refuge. Calmann-Levy, Paris

Efron R (1956) The effect of olfactory stimulus in arresting uncinate fits. Brain 79:267–281

Efron R (1957) The conditioned inhibition of uncinate fits. Brain 80:251–262

Ellickson JL (1983) Representational systems and eye movements in an interview. Journal of Counseling Psychology 30(3):339–345

Freeman W (2001) Neurodynamics. Springer Verlag, Berlin

Galin D, Orstein R (1974) Individual differences in cognitive style and reflective eye movements. Neuropsychologia 12:76–397

Gallagher S (2000) Philosophical conceptions of the self: implications for cognitive science. Trends Cogn Sci 3(1):14–21

Gastaut H (1954) The epilepsies, electro-clinical correlations. Thomas, CC, Springfield, IL

Gendlin E (1962/1997) Experiencing and the creation of meaning. Northwestern University Press, Evanston, IL

Gendlin E (1996) Focusing-oriented psychotherapy. The Guilford Press, New York

Grinder J, Delozier J, Bandler R (1977) Patterns of the Hypnotic Techniques of Milton Erickson, vol. 2. Capitola, CA, Meta Publications

Husserl E (1913/1977) Ideen zu einer reinen phaenomenologie und phaenomenologischen philosophie. Kluwer, The Hague, Netherlands

Kinsbourne M (1972) Eye and head turning indicates cerebral lateralization. Science 179:539–541

Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. Hum Brain Mapp 8:194–208

Le Van Quyen M (2003) Disentangling the dynamic core: a research program for neurodynamics at the large scale. Biol Res 36:67–88

Le Van Quyen M, Petitmengin C (2002) Neuronal dynamics and conscious experience: an example of reciprocal causation before epileptic seizure. Phenomenol Cogn Sci 1:169–180

Le Van Quyen M, Martinerie J, Navarro V, Baulac M, Varela F (2001) Characterizing the neuro-dynamical changes prior to seizures. J Clin Neurophysiol 18:191–208

Le Van Quyen M, Navarro V, Martinerie J, Baulac M, Varela F (2003) Toward a neurodynamical understanding of ictogenesis. Epilepsia 44:30–43

Le Van Quyen M, Soss J, Navarro V, Robertson R, Chavez M, Baulac M, Martinerie J (2005) Preictal state identification by synchronization changes in long-term intracranial EEG recordings. J Clin Neurophysiol 116:559–568

Levine J (1983) Materialim and qualia: the explanatory gap. Pac Philos Q 64:354–361

Lutz A (2002) Toward a neurophenomenology of generative passages: a first empirical case study. Phenomenol Cogn Sci 1:133–167

Lutz A, Lachaux JP, Martinerie J, Varela J (2002) Guiding the study of brain dynamics using first person data: synchrony patterns correlate with on-going conscious states during a simple visual task. Proc Natl Acad Sci USA 99:1586–1591

Lutz A, Greischar L, Rawlings N, Ricard M, Davidson R (2004) Long-term meditators self-induce high amplitude gamma synchrony during mental practice. PNAS 101(46):16369–16373

Lutz A, Thompson E (2003) Neurophenomenology. Integrating subjective experience and brain dynamics in the neuroscience of consciousness. J Conscious Stud 9–10:31–52

Marchand L, de Ajuriaguerra J (1948) Les épilepsies. Desclée de Brouwer, Paris

Merleau-Ponty M (1964) L'œil et l'esprit. Gallimard, Paris

Mormann F, Lehnertz K, David P, Elger CE (2000) Mean phase coherence as a measure for phase synchronization and its application to the EEG of epileptic patients. Physica D 144:358–369

Mormann F, Andrzejak RG, Kreuz T, Rieke C, David P, Elger CE, Lehnertz K (2003a) Automated detection of a preseizure state based on a decrease in synchronization in intracranial electroencephalogram recordings from epilepsy patients. Phys Rev E 67. doi: 021912

Mormann F, Kreuz T, Andrzejak RG, David P, Elger CE, Lehnertz K (2003b) Epileptic seizures are preceded by a decrease in synchronization. Epilepsy Res 53:173–185

Nagel T (1970) What is it like to be a bat? Philos Rev 79:394–403

Navarro V, Adam C, Petitmengin C, Baulac M (2006) Toothbrush-Thinking Seizures. Epilepsia 47:1971–1973

Petitmengin C (2001) L'expérience intuitive. L'Harmattan, Paris

Petitmengin C (2005) Un exemple de recherche neuro-phénoménologique : l'anticipation des crises d'épilepsie. Intellectica 40:63–89

Petitmengin C (2006a) L'énaction comme expérience vécue. Intellectica 43:85–92

Petitmengin C (2006b) Describing one's subjective experience in the second person. An interview method for the science of consciousness. Phenomenol Cogn Sci 5:229–269

Petitmengin C (2007) Towards the source of thoughts. The gestural and transmodal dimension of lived experience. J Conscious Stud 14(3):54–82

Petitmengin C (2009) Ten years of viewing from within. Imprint Academic, London (in press)

Petitmengin C, Navarro V, Baulac M (2006) Seizure anticipation: are neurophenomenological approaches able to detect preictal symptoms? Epilepsy Behav 9(2):298–306

Petitmengin C, Navarro V, Le Van Quyen M (2007) Anticipating seizure. Pre-reflective experience at the center of neurophenomenology. Conscious Cogn 16(3):746–764

Petitmengin C, Bitbol M (2009) The reliability and validity of first person descriptions. In: Petitmengin (ed) Ten years of viewing from within. Imprint Academic, London (in press)

Petitmengin-Peugeot C (1999) The intuitive experience. In: Varela FJ, Shear J (eds) The view from within. First-person approaches to the study of consciousness. Imprint Academic, London, pp 43–77, 1999a

Rajna P, Clemens B, Csibri E, Geregely A, Gottschal M, György I, Horvath A, Horvath F, Mezöfi L, Velkey I, Veres J, Wagner E (1997) Hungarian multicentre epidemiologic study of the warning and initial symptoms (prodrome, aura) of epileptic seizures. Seizure 6:361–368

Reiter J, Andrews D (2000) A neurobehavioral approach for treatment of complex partial epilepsy efficacy. Seizure 9:198–203

Ricœur P (1949) Philosophie de la volonté. Le volontaire et l'involontaire. Aubier, Paris

Roy JM, Petitot J, Pachoud B, Varela F (2002) Combler le déficit : introduction à la naturalisation de la phénoménologie. In: Petitot J, Varela F, Pachoud B, Roy JM (eds) Naturaliser la phénoménologie. CNRS Éditions, Paris, pp 1–100

Sartre JP (1936) L'imagination. Presses Universitaires de France, Paris

Sartre JP (1938) Esquisse d'une théorie des émotions. Hermann, Paris

Schmid-Schönbein C (1998) Improvement of seizure control by psychological methods in patients with intractable epilepsies. Seizure 7:261–270

Schulze-Bohage A, Kurt C, Carius A, Steinhoff B, Mayer T (2006) Seizure anticipation by patients with focal and generalized epilepsy: a multicentre assessment of premonitory symptoms. Epilepsy Res 70:83–88

Soulayrol R (1999) L'enfant foudroyé. Comprendre l'enfant épileptique. Odile Jacob, Paris

Thompson E, Varela F (2001) Radical embodiment: neuronal dynamics and consciousness. Trends Cogn Sci 5:418–425

Stern D (1985) The interpersonal world of the infant. Basic Books, New York

Varela F (1995) Resonant cell assemblies: a new approach to cognitive functions and neural synchrony. Biol Res 28:81–95

Varela F (1996) Neurophenomenology: a methodological remedy for the hard problem. J Conscious Stud 3:330–335

Varela F (1997) The naturalization of phenomenology as the transcendance of nature: searching for generative mutual constraints. Alter: Rev Phénoménol 5:355–385

Varela F (1999a) Steps to a science of interbeing: unfolding the Dharma implicit in modern cognitive science. In: Bachelor S, Claxton G, Watson G (eds) The psychology of awakening: Buddhism, science and our day to day lives. Rider/Random House, New York, pp 71–89

Varela F (1999b) Quatre phares pour l'avenir des sciences cognitives. Théorie Littérature Enseignement 17: Dynamique et cognition: nouvelles approches

Varela FJ, Shear J (eds) (1999a) The view from within. First-person approaches to the study of consciousness. Imprint Academic, London

Varela FJ, Shear J (1999b) First-person methodologies: What, why, how? In: Varela FJ, Shear J (eds) The view from within. First-person approaches to the study of consciousness. Imprint Academic, London, pp 1–14, 1999a

Varela FJ, Thompson E, Rosch E (1993) The embodied mind: cognitive science and human experience. MIT Press, Cambridge

Varela F, Lachaux JP, Rodriguez E, Martinerie J (2001) The brain web: phase synchronisation and large scale integration. Nat Rev Neurosci 2:229–239

Vermersch P (1994/2003) L'entretien d'explicitation. ESF, Paris

Wolf P (1997) Behavioral therapy. In: Engel J, Pedley TA (eds) Epilepsy: a comprehensive textbook. Lippincott-Raven, Philadelphia, PA, pp 1359–1364

How Unconscious is Subliminal Perception?

Morten Overgaard and Bert Timmermans

Ever since the concept of the "unconscious" defragmented one of the last nuclei of certitude in the western world, one of the great mysteries that common knowledge has come to associate with psychology is the idea that much of what goes on in our lives occurs outside of awareness. In other words, what we experience is but the tip of the iceberg: massive amounts of information are processed, have an influence on our behaviour, all the while going unnoticed. But is this the case? It seems indeed that we are not always able to retrace the associative causal chains that lead to a thought or action (though the degree to which this is the case is still being debated) – but according to the way we consciously experience the world, it seems to be both immediate and complete to us. Apart from the fact that we may not be able to perceive the entirety of the external world due to information processing limitations inherent to our physical system, what seems more unsettling is that we may not even be able to experience all that we perceive! Is there some limitation of our own internal system that keeps not only the world, but the way this world is represented in us fundamentally unknowable to ourselves?

Many psychological experiments have attempted to prove the existence of unconscious perceptual processes by demonstrating that stimuli are perceived when subjects are not consciously aware of the stimuli. The basic strategy followed in these studies is to establish conditions under which conscious experience does not occur and then to demonstrate that stimuli can nevertheless be perceived under these conditions. The success of these studies depends crucially on two factors: first, the acceptability of the method used to establish the absence of conscious perception, and second, the method of assessing that the not consciously experienced stimulus is indeed unconsciously perceived. In the earliest studies, inferences concerning the absence of awareness were based on subjects' introspective reports.

M. Overgaard
CNRU, Hammel Neurorehabilitation and Research Center, Aarhus University Hospital,
Voldbyvej 15, 8450, Hammel, Denmark

B. Timmermans
Consciousness, Cognition, & Computation, Universite Libre de Bruxelles, CP 191,
Av FD Roosevelt 50, B1050, Brussels, Belgium

In general, if the subjects' statements indicated an absence of relevant conscious perceptual experiences, it was assumed that the subjects were in fact unaware of the stimuli. In more recent studies, the absence of relevant conscious experiences has been defined in terms of behavioral measures that indicate an inability to discriminate between alternative stimuli.

Studies based on both types of measures have not led to completely convincing results because it is always possible to question whether the measure of conscious perception was successful in guaranteeing a complete absence of all relevant conscious experiences, as absence of evidence is no evidence of absence. Likewise, it is difficult to assess whether any residual discriminatory acuity does not result from unconscious knowledge, nor is it easy to know how both types of knowledge (conscious and unconscious) contribute to behavioral measures that are supposed to reflect pure unconscious perception. Apart from the fact that on top of all this it is conceptually difficult to assess the degree to which any fragmentary fine-grained piece of information is actually conscious (is "having a hunch" conscious?), we will see that some solutions have been put forward to this problem, such as d', Merikle's qualitative differences approach, or Jacoby's process dissociation procedure, neither of which is entirely satisfactory when our aim is to capture the phenomenology of conscious experience.

The aim of this chapter is both to provide a general overview of the past century's pendulum-movement towards, away from, and again towards subliminal perception and unconscious processes, and how this is inseparable from the way one measures conscious awareness. We will put forward several arguments in favour of (re-)adopting a subjective, introspective approach, which, while having its limitations, allows not only for capturing the richness of phenomenal experience, but also may question the very existence of subliminal perception.

Short Biographical Sketch of Subliminal Perception

The First Subliminal Wave – Perception is Not Just Stimulus-Related

The idea that psychologists can reveal the existence of unconscious cognitive processes experimentally by demonstrating that stimuli can be perceived without the subject's conscious experience has been part of experimental psychology since its very beginning. Yet, the question of so-called subliminal perception is often presented as one of the most controversial issues.

Throughout the historical development of the view on subliminal perception, the definition of the concept itself has changed, marked again by a pendulum-movement with regards to the relative independence of conscious experience and perception. In the very beginning of perception studies, the concept of "subliminal" referred to characteristics of a *stimulus* where its energy level was too weak for conscious perception. In other words, conscious experience was more or less considered

equivalent to perception: when the stimulus was weak and conscious experience decreased, this meant that perception diminished to the same degree.

Peirce and Jastrow (1885) were the first who went *against* the notion that at some low energy level, stimuli simply cannot be picked up by the sensory organs, and were probably the first to empirically demonstrate subliminal perception, conceptualised as perception in the absence of conscious experience. Using themselves as subjects (criteria for good scientific practice were less strict at the time) Peirce and Jastrow found that they could make accurate forced-choice judgments of relative weight or brightness of objects, even when they reported no confidence in their own judgments. To describe experiences, Peirce and Jastrow agreed on an intuitive scale:

Denoted absence of any preference for one answer over its opposite, so that it seemed nonsensical to answer at all.

1. Denoted a distinct leaning to one alternative.
2. Denoted some little confidence of being right.
3. Denoted as strong a confidence as one would have about such sensations

A few years later, Boris Sidis (1898) presented subjects with small cardboard cards, each containing a single printed letter or digit. The distance between the subjects and the cards was such that the subjects often complained that all they could see on each card was a dim, blurred spot or nothing at all (remember that this was well before tachitoscopic presentation and masking). Based upon this, Sidis assumed that the subjects were unaware of perceiving either digits or letters. However, when he used a second measure, forced-choice guessing, he discovered that his subjects were able to guess the category of the card (digit or letter). Furthermore, he discovered that the subjects were better than chance at guessing the precise identity of the card. Thus, Sidis described *dissociation* between two measures of perception, in themselves independent of the stimulus characteristics. The *subjective, verbal measure* from the subjects suggested that they did not have a visual experience of the critical stimuli, while the *objective, behavioural measures* from forced-choice guessing suggested that the subjects indeed had perceived the stimuli. Sidis suggested that the results indicate: "(…) the presence within us of a secondary subwaking self that perceives things which the primary waking self is unable to get at" (1898, p. 171). Hence, the findings also provoked a theoretical discussion about the relationship between perception and consciousness: is there one perception plus or minus consciousness, or are different conscious experiences still reflecting differences in perception? And how does behaviour relate to what is consciously or unconsciously perceived? As we will see, many of these issues still remain unresolved to this day.

Other contemporary psychologists applied methodologies similar to Sidis and his results are well established (Stroh et al. 1908). As such, there was very little debate at the time about the reliability of the finding that subjects report statistically correct about the shapes of objects even under conditions where the object was hard to see and when they reported no conscious vision. However, due to debates of a more paradigmatic nature in all fields of scientific psychology, there was

an increasing worry about the general validity of such introspective reports. As Merikle and Daneman (1998), Merikle (1984), it is difficult to know which subjective criteria, subjects use when reporting. This may be of special difficulty when subjects report not seeing anything. In such cases, subjects may report not seeing anything when, in fact, they are just not able verbally to specify the nature of a conscious sensation. Instead, a number of papers argued that objective methods must replace the reports.

Discredit of Subjective Methods – Towards Objective Measures

As we saw, in the early years, subliminal perception studies focused on the apparent dissociation between perception, as measured objectively by behavior, and awareness, as measured by subjective reports. However, this discovery sparked new controversies that are still very much alive, since, as noted by Erdelyi (1985), every attempt to demonstrate subliminal effect has to some degree relied on a version of this dissociation paradigm. The crucial controversies revolve around lacking agreement about what constitutes a measure of awareness, or consciousness. To what degree *can* we rely on subjective reports, and if not at all, then *what* can be a good measure, especially since we need a measure to *rule something out*, so that its outcome should ideally be zero? To frame the problem more clearly, an ideal measure of conscious knowledge should not only be *exhaustive* but also *exclusive*. By exhaustive, we mean that the measure or task should be sensitive to every bit of conscious knowledge, so that we may avoid erroneously describing behavior as resulting from subliminal perception (an issue to which we will return later). By exclusive, we mean that results on such a task should not be prone to be influenced by unconscious knowledge, so that we may avoid any test that, rather than measuring awareness, is just another behavioral measure of subliminally perceived stimuli.

Exhaustiveness of Subjective Methods

Eriksen (1956, 1960; see also Goldiamond 1958) argued that the issue debated by the early experimental psychologists was not just "subliminal perception" in the sense of "perception of very weak stimuli" but rather perception in the total lack of conscious awareness of the perceived object. He argued that verbal reports were insufficient in this regard as they do not, or may not, represent an exhaustive measure of consciousness. As indicated above, the problem especially arises in cases where subjects report no conscious awareness where in fact there may be some vague experience or sensation, which is hard to express verbally. Eriksen argued that *forced-choice identification tasks*, also known as discrimination tasks, provide a measure of discriminability (often just referred to as d' in experimental literature).

Thus, subjective reports might reflect a participant's response criterion to one specific conscious process rather than being indicative of the boundary between

conscious and unconscious experience. Discriminability on the other hand, is held to be independent of such a bias according to Signal Detection Theory (Green and Swets 1966; Macmillan and Creelman 1991). Subsequently, forced-choice identification tasks have come to be known as 'objective measures of awareness'. This has led for instance Merikle (Cheesman and Merikle 1984, 1986; Reingold and Merikle 1988, 1990) to propose the *subjective threshold* model. This model states that when we measure any subliminal perception, that is, above chance discrimination performance in the absence of awareness, this is merely due to participants' incomplete reports. Obviously, this very definition makes subliminal perception a priori impossible, since the phenomenon is understood as visual abilities (e.g. discrimination) in the absence of consciousness. Were consciousness to be operationalized as one such "visual ability", clearly, one could never find "visual ability in the absence of consciousness". In other words, d' may be an exhaustive measure, but not an exclusive one, as zero discriminability may in fact rule out any unconscious perception or knowledge as well, making it impossible to be certain of what is in fact being measured, if anything at all (Dixon 1971, 1981; Jacoby et al. 1992; Merikle and Daneman 1998).

More recently, Greenwald and colleagues (Draine and Greenwald 1998; Greenwald et al. 1996) have proposed the existence of an additional threshold in their *objective threshold/rapid decay* model, in that they assume that objective threshold effects are real, but very short lived, whereas subjective threshold effects, as reported by Merikle, are probably weak conscious effects, which participants fail to report. Even more recently, Snodgrass and colleagues (Snodgrass et al. 2004a) proposed a third, *objective threshold/strategic* model, which also proposes that objective threshold effects are genuine. They adopt a dual process view of perception. In their view, every process has relatively independent conscious and unconscious components, and as a consequence, not only conscious perception effects gradually become stronger with increased stimulus intensity, but also unconscious perception effects become *weaker* as these conscious effects override them

An additional problem for discrimination tasks is that, even if they were ever demonstrated to be exhaustive and exclusive, however exactly this should be done, the next step would be to demonstrate actual null sensitivity, following Eriksen (1960) and all modern interpretations of subliminal perception. Many studies reporting under chance discrimination (Balota 1983; Marcel 1983) have not demonstrated a total lack of conscious experience. According to MacMillan (1986), the number of trials necessary to show null from low-level sensitivity to huge – $d' = 0$ vs $d' = 0.5$ – is 140 trials (Reingold 2004), which makes the documentation of no experience experimentally very difficult.

Exclusiveness of Subjective Methods

Apart from being insufficient, or not exhaustive, subjective reports have been under attack for being unreliable, or not exclusive. In other words, what participants report they saw may not be merely indicative of their conscious experience, but instead may also be prone to unconscious influences. In their seminal paper, Nisbett and Wilson (1977) reviewed a large number of studies in social psychology,

all demonstrating a surprising lack of ability in subjects to account for the cause of or processes involved in their own behavior. For instance, they describe a study where subjects are asked to report their preference for two presented items. Subjects showed a systematic tendency to prefer items presented to the right of them, although, when asked about the reasons for their choice, they simply reported that they made their choice based on which items they found the prettiest, and not on which item was located to the right. Based on their review, Nisbett and Wilson argued that "there is by now enough evidence discrediting introspective reports to allow us to ignore any argument based on introspection" (p. 255).

It could be argued, however, that this conclusion misconstrues the evidence. Subjects giving an introspective report about liking objects presented to the right for some other reason than the object's location in space may be giving a perfectly good and scientifically usable report of what they experienced. Nisbett and Wilson correctly rejected introspection as a methodology to learn about (some aspects of) choice and decision making, as the behavioral data suggested a very different explanation from the one that subjects themselves reported. Another interpretation of the results could be, however, that in some unknown (but probably vast) number of situations, people do not have introspective access to their own cognitive processes, or to the causal chain of factors that have influenced a certain action. However, not surprisingly, they still have some experience and interpretation of their own actions. Thus, a conflict in data between subjective report and behavior could be interpreted to show that the subject's experience differs from what can be analyzed from his or her behavior, and, thus, it does not automatically follow that the introspective report is invalid. Indeed, it has been argued (Schachter and Singer 1962) that people learn most of their self-knowledge observing their own behavior, which might not always lead to the correct conclusions, something known as the attribution error. For instance (Dutton and Aron 1974) showed that men, when given a questionnaire by a pretty female experimenter *after* having crossed a precipitous bridge were more likely to return the questionnaire than men whom she had given the questionnaire to on the street. The authors suggested that the men misattributed their arousal to feelings for the female experimenter.

Ericsson and Simon (1984) were historical opponents to Nisbett and Wilson, arguing that highly valuable information lies in subjective reports, and that they cannot be replaced by any other source of data. Instead, Ericsson and Simon suggest that more elaborate methods for the use of subjective reports should be developed although this appeal never set a strong mark on experimental psychology.

Discredit of Subjective and Objective Methods – Focus on Processing

Absence of Evidence

In 1988, Rheingold and Merikle restated the previously mentioned points of criticism, stressing that, apart from possibly being not exclusive and thus ruling out

the possibility of finding any subliminal effects, the most fundamental problem for the idea of the objective threshold remains its exhaustiveness. Specifically, the fact that its claim to fame as a consciousness measure rests on the *assumption* that it *is* exhaustive, makes it something which is *de facto* impossible to prove, since, as stated at the start of this chapter, absence of evidence can never be evidence of absence. For instance, above chance performance in a semantic priming task with a complete zero discriminability does not guarantee anything if one assumes that different measures may tap into different aspects of task-relevant knowledge. Therefore, Reingold and Merikle (1988) suggested comparisons of the *relative sensitivity* of different types of assessments of perception and memory as one method to identify such differences. They proposed that an important distinction between tasks concerns whether a task provides a direct or an indirect index of a particular stimulus discrimination (see also Richardson-Klavehn and Bjork 1988). Whenever the indirect measure yields better performance, unconscious processes play a role. This approach effectively avoids the exhaustiveness problem, as no awareness needs to be shown as zero.

Later this relative sensitivity approach was replaced when Phil Merikle (1992; Cheesman and Merikle 1986; Merikle et al. 1995) suggested looking for a *qualitative difference in processing* of stimuli presented under different perceptual conditions, as an alternative research strategy. The idea is that conscious and unconscious perceptual processes differ "qualitatively", meaning that they are associated with different behavioral outcomes. Tasks in which subjects are explicitly instructed to perform the memory or perceptual discrimination of interest are defined as direct measures of memory or perception. In contrast, if the instructions given to subjects do not make any reference to the discrimination of interest, then such tasks are defined as indirect measures. Unconscious processes are implicated, they say, whenever an indirect measure shows greater absolute sensitivity than a comparable direct measure to a particular stimulus discrimination. This is the case because the assumption rules out the possibility that superior performance on the indirect task is attributable to conscious task relevant information. Therefore, whenever an indirect measure indicates greater sensitivity than a comparable direct measure, it must reflect a greater sensitivity of the indirect measure to unconscious, task relevant information. In this context, Jacoby (1991) proposed the *process dissociation procedure*, which aims at separating automatic from intentional uses of memory by means of an inclusion and an exclusion task. Typically, in the inclusion task participants are asked to reproduce what they can recall from memory, their correct recall reflecting both conscious and unconscious memory processes (the direct measure). In the exclusion task however, participants are asked to *avoid* reproducing what they can recall (the indirect measure). Under these instructions, above chance correct recall despite the instruction not to do so, reflects only unconscious memory processes. In this way, the problem of showing exhaustiveness by means of null sensitivity is avoided.

One illustration of the search for qualitative differences is a *semantic priming* study reported by Marcel (1980). In this experiment, subjects were presented with a series of three, successive letter strings. On critical trials, the second letter string was a polysemous word (e.g., PALM). In the congruent condition, the first and the third letter

strings were words related to the same meaning of the polysemous word (e.g., HAND, PALM, WRIST), while in the incongruent condition, these words were related to different meanings of the polysemous word (e.g., TREE, PALM, WRIST). The task for the subjects was to decide whether the third and final letter string was a word or a nonword (i.e., a nonsense string of letters). When the polysemous words were clearly visible, Marcel found that the decisions concerning the third letter string were only facilitated (i.e., faster) on congruent trials. In contrast, when the visibility of the polysemous word was severely degraded, the results indicated facilitation on both congruent and incongruent trials. Marcel suggested that these results indicate that conscious perception involves selection of the one meaning consistent with the context (i.e., the first word), whereas unconscious perception is not constrained by context.

The qualitative differences approach has been widely accepted and may be said to dominate current literature on the subject together with methods that only employ objective measures. In particular, it assumes that any qualitative difference found can be explained simply by differential contributions of conscious and unconscious influences on behavior. Nonetheless, Merikle and colleague's approach can be criticized in the same way as forced-choice identification tasks as a measure of consciousness. As Holender (1986; see also Holender and Duscherer 2004) has noted, qualitative differences in processing may arise over thresholds other than the threshold for consciousness. For instance, the exclusion measure may just fail to be exclusive. It could be that, like the "objective" discrimination task, exclusion reflects a criterion bias. Failure to exclude could reflect participants' uncertainty about whether they saw that specific stimulus, leading them to report it under exclusion conditions.

The Development of the Definition of Subliminal Perception

Following Marcel (1983a), most studies have evolved around perceiving the semantic contents of words. Also, with the invention of the tachistoscope (and later computers) as employed by Marcel, the use of weak stimuli changed to brief stimuli. With the use of backward masking techniques, the concept of "subliminal" really made little sense. Here, a stimulus is presented with a strength and duration typically sufficient to allow for conscious perception. Yet with the application of a visual mask after (or prior) to the stimulus, the stimulus seems not to be consciously perceived. Even newer methods take advantage of the "attentional blink" phenomenon: When a sequence of visual stimuli is presented in rapid succession at the same spatial location (rapid serial visual presentation, or rsvp), a subject would typically fail to detect a second stimulus if the first was perceived (Raymond et al. 1992).

The Second Subliminal Wave – What About Subjective Reports?

As we have seen, the acceptance of subliminal perception as a real phenomenon has varied throughout the history of psychology. During the last decades, and

especially since the late 1980s, an increasing amount of evidence has suggested that unconscious processes are not limited to phenomena like subliminal discrimination abilities in the sense of an ability to make correct guesses about object features, such as the early studies showed. A number of studies have now suggested that we can also have subliminal semantic processing, subliminal emotion, subliminal social comparisons or subliminal thought.

In experiments conducted by Murphy and Zajonc (1993), subjects were shown a clearly-visible, Chinese ideograph on each of a series of trials. The subjects were asked to indicate on a five-point scale whether they thought each ideograph represented a "good" or a "bad" concept. The critical aspect of the experiment concerned what happened immediately before each ideograph was presented. For one group of subjects, the presentation of each ideograph was preceded by a picture of a human face that expressed either happiness (a smile) or anger (a scowl). For this group of subjects, each face was presented for such a brief duration (4 ms) that no subject reported awareness of the faces. For the second group of subjects, the same ideographs and faces were presented, but the duration of each face (i.e., 1,000 ms) was sufficiently long so that all subjects reported awareness of the faces. The subjects in this second group were told to ignore the faces and to concentrate solely on rating the ideographs. The important result found by Murphy and Zajonc is that only the briefly-presented, unconsciously perceived faces influenced the subjects' ratings of the ideographs. When the subjects were unaware of the faces, they were more likely to rate an ideograph as representing a "good" concept if it was preceded by a smiling face and they were more likely to rate an ideograph as representing a "bad" concept if it was preceded by a scowling face. In contrast, when the faces were clearly visible and therefore consciously perceived, the faces had little or no influence on the subjects' ratings of the ideographs. Thus, the subjects were able to ignore consciously perceived faces and not let these faces influence their ratings of the ideographs. However, when the subjects were unaware of the faces, the emotion expressed by the faces colored their judgments of the ideographs, a finding which is in line with Snodgrass et al. (2004)'s objective threshold/strategic model. Unconscious emotional reactions, of course, are one thing.

Much more controversial were the illustrations of unconscious semantical processing produced by Marcel in an article in Cognitive Psychology (1983). In the described experiment, subjects made judgments of graphic and semantic similarity for words that presumably could not be detected. One major finding was that at low exposure rates semantic judgments were superior to graphic judgments, which in turn were more probable than correct "presence or absence" judgments. This effect was related to subjects who "adopted a 'passive' attitude" and chose the words that "felt" right. Marcel's experiment illustrates the three major components of any demonstration of unconscious perception, based on the dissociation paradigm. First, he used stimulus detection as the measure of conscious perception. Second, he established the experimental conditions that made it impossible for the participants to discriminate at a better than chance level of accuracy between the presence or absence of the stimuli. Third, he used a measure of semantic priming that was based on reaction time to show that despite the participants' inability to detect the

stimuli, it was still possible for the stimuli to prime subsequent decisions regarding other stimuli.

As a whole, Marcel's study was an important part of re-establishing subliminal perception as an accepted phenomenon in experimental psychology. Furthermore, the finding that something so "advanced" as semantic processes were possible without consciousness was certainly part of forming the widespread idea among psychologists today that most of our mental life is in fact unconscious. However, recently Kouider and Dupoux (2004) suggested that in fact semantic priming effects could already result from partial awareness of certain letters of the prime, which trigger an illusory reconstruction of the target stimulus. For instance, they found subliminal semantic priming effects for words, such as "BLUE", as well as for their scrambled versions, such as "BELU", when participants could only discriminate some letters, but *not* when they were aware the word, *nor* when they weren't aware of any of the letters, in a completely subliminal condition.

During the past decade or so, we have seen a rekindled interest in trying to show the existence of subliminal effects in domains such as perception, memory, and action – very often inspired by the search for the neural correlates of consciousness (NCC). Hereby, the focus lies either on identifying (a) equal levels of performance, accompanied by different degrees of awareness (e.g., blindsight), (b) changes in performance unaccompanied by changes in awareness (e.g., implicit learning), and (c) changes in awareness despite perception remaining constant (e.g., binocular rivalry). The favorite example of subliminal abilities among many researchers today is for instance the phenomenon of "blindsight". Blindsight refers to the impressive discovery that at least some patients with lesion to the primary visual cortex have preserved visual functions such as perception of movement direction (Weiskrantz et al. 1995), target detection (Pöppel et al. 1973) and spatial summation (Leh et al. 2006) even though they report to be fully blind in a part of the visual field corresponding to the location of the injury. As such, blindsight should be considered "less interesting" than subliminal perception in healthy subjects, as the phenomenon has so far only been studied in a few patients. However, in those patients, blindsight has shown so consistent and persuasive as example of an almost unbelievable discrepancy between subjective report and behavioural reactions (such as the ability to discriminate) that many researchers see it as the primary source of evidence for subliminal processing. In 1989, however, Weiskrantz and co-workers found evidence to argue that blindsight should be subdivided into two "types" – type 1 and type 2. Type 1 blindsight patients are characterized, as above described, by preserved visual functions despite of verbal reports of having no visual experiences. Type 2 blindsight patients report seeing after-images or "shadows" when presented with stimuli.

Also, in the decision-making and social domain, the past ten years have seen a real boom of "unconscious effects" research. (Dijksterhuis 2004; Nordgren and Dijksterhuis 2005) investigated how subjects made choices under different conditions in experiments where subjects were asked about their preference when hypothetically buying an apartment. In the experiments, some people were not given the opportunity to think at all before choosing between alternatives.

Others were allowed to think for a while before choosing, and yet others were distracted for a while before choosing and thus could only engage in what Dijksterhuis and Nordgren refer to as unconscious thought (Dijksterhuis and Nordgren 2006). In one experiment, participants were given information about four hypothetical apartments in their home city, Amsterdam. Each apartment was described by 12 different aspects (e.g., apt. A is rather sizable, apt. C is in a nice area, apt. D has a very unfriendly landlord) adding up to a total of 48 pieces of information. One of the four apartments was made more desirable than the others (it had predominantly positive aspects), whereas a second one was made undesirable (it had predominantly negative aspects). The two remaining apartments were more neutral. After participants had read the information, they were asked to evaluate each apartment. The so-called unconscious thinkers performed significantly better than those with the opportunity to reflect and the immediate choosers. In fact, the subjects who were allowed to reflect did not indicate greater liking of the desirable apartment compared to the undesirable apartment. Dijksterhuis' results are in the line of those of Nisbett and Wilson (1977), and indeed his main conclusion is that people can engage in unconscious thought processes, or, put otherwise, that people do not have access to a number of thought processes. In the same line, various experiments have shown that it is possible to subliminally influence people's self-esteem (Baldwin 1994; Dijksterhuis 2004; Timmermans et al. 2008; for an overview, see Greenwald and Banaji 1995). Recent experiments on priming of unconscious processes has even shown that one can positively influence people's cleanliness by means of having cleaning product scent in the experimental room (Holland et al. 2005), that, when unconsciously primed with words related to old age, people tend to walk slower than those who haven't been primed (Bargh et al. 1996), and a whole range of other unconscious social phenomena (see Bargh 2007, for an overview).

Collectively, such studies have served to draw a picture of human beings as primarily influenced by unconscious processes and who function more efficiently and, perhaps even more rationally unconsciously than they do consciously. There is no doubt that the modern-day understanding of subliminal processes is significantly different from the understanding of "below threshold" sensation among the late nineteenth century psychologists.

Objective and Subjective Measures, and What They Can Tell Us About Subliminal

Why We Need More Than Objective Measures

Let us briefly summarize the issues discussed so far. Demonstrating the existence of subliminal processing depends crucially on the measurement accuracy of conscious experience. Such a measure must be both exhaustive and exclusive.

We have seen however, that none of the attempts to overcome exhaustiveness and exclusiveness issues are completely successful. Also, and perhaps more crucial, these attempts have gradually done away with any actual description of the richness of the phenomenology of conscious content. The issue at stake is simply: what does a person experience and how does he experience it?

It is important to note that this conclusion implies questions about both the nature (what is consciously perceived?) and the degrees of the knowledge (how much of it is perceived?), two aspects that are obviously interconnected, because if one consciously perceives part of a stimulus (e.g., a triangle), this will often have implications for what is perceived (just one oblique line, i.e. one side of the triangle). Consequently, it is crucial to establish the degree to which that conscious knowledge is in a clear way related to the stimulus; in other words: do we tap into the relevant kind of knowledge? If someone is conscious of an oblique line, to what degree can we say that she has been conscious of information relevant for the task? Will this information influence her behavior? This may not be the case; on the other hand, it can be that the participant doesn't know what knowledge helped her with the task, simply because the knowledge is conceptually different from anything experienced, either because the stimulus was subtreshold and they lack consistent meaningful knowledge of the stimulus as a whole, or because they lack insight in how their decision can be causally traced back to a stimulus. These aspects may be of slightly less importance when we want to demonstrate subliminal perception, and thus *eliminate* consciousness. However, they are extremely important when we want to know anything about the nature or the content of consciousness, so when we want to *explore* consciousness.

The crucial issue here is that the attempt to "replace" subjective methods in any direct sense with objective methods is in itself a very problematic enterprise. Arguing, say, that some objective method like forced-choice discrimination lends a "more direct" insight into the contents of consciousness rests upon circularity (Overgaard 2006). There may be fixed contingent relations between certain responses and experience, so that the presence of such gives us right to claim that a subject has a certain experience. However, finding the correct objective measures is impossible without making use of subjective data, e.g. an introspective report. That is, associating a certain report such as a correct identification with consciousness is only possible with empirical evidence, i.e. a correlation between the response and the relevant conscious state. Since the conscious state cannot in itself be observed from the outside, the use of an introspective report about the relevant state seems the only possible methodology. Accordingly, no other kind of response can be a more reliable indication of a given conscious state than the introspective report. This conclusion logically follows from the fact that the response is associated with the conscious state only by way of its correlation with the introspective report. Thus, for instance, if we are looking for NCC that correlate with phenomenal richness of conscious experience, we need a tool that can tell us something about this richness.

Is there, at all, a subjective side to subliminal perception? This question is rarely asked and may seem somewhat strange as fully unconscious states hardly can be said to have a "subjective side". However, the subjects participating in experiments

investigating subliminal perception are not fully unconscious. A conscious subject has conscious states and, therefore, she has access to knowledge regarding the presence or absence of her own conscious states. Accordingly, research in subliminal perception must be carried out in agreement with introspection. That is, it would not be an acceptable case of subliminal perception, if a research subject insisted that she did in fact have a conscious experience. Therefore, subliminal perception must be studied with complementary methods, and no real conclusions regarding the nature or the very existence of unconscious processes can be made without some measure based on introspection.

With regards to the claim that subjective reports (hence introspective reports) are not exhaustive, it is an open question whether to interpret this criticism to mean that subjective reports should not be trusted (that they are not exhaustive as a matter of principle) or whether it means that subjective reports must be methodologically refined in order to become exhaustive.

Why We Need a Continuous Subjective Measure – The Perceptual Awareness Scale (PAS)

Ramsøy and Overgaard (2004) tried a different approach to introspective information on conscious and unconscious processes. Subjects were here asked to create their own categories for subjective reports during long training sessions. They were asked what they were shown and how they experienced stimuli in terms of clarity. Here, stimuli were simple visual figures (triangles, circles or squares) presented in one of three possible colours (blue, green or red). In the study, subjects conformed to a four-point scale categorised as "not seen", "weak glimpse" (meaning "something was there but I had no idea what it was"), "almost clear image" (meaning "I think I know what was shown") and "clear image". When subjects tried to use more than four categories in the scale, they found it confusing and quickly abandoned the extra categories. In the experiment, after the category-generating training process, subjects found the categories easy to use, and in free reports, they explained that the categories seemed very straightforward.

Ramsøy and Overgaard showed that in an experimental design where one should expect to find subliminal perception, there was in fact none – at least not in the understanding of Eriksen and all contemporary authors. In a later study (Overgaard et al. 2006), two different report methods were compared directly to investigate subliminal perception.

In this experiment, subjects were presented with a series of oriented-element textures. The "target" consisted of four orthogonal elements in one of the four possible quadrants of the screen. The shortest stimulus duration was 12 ms, and multiples of 12 ms for longer durations were determined by the monitor refresh rate. A mask, displayed for 500 ms before the target stimulus and 800 ms after, consisted of an image of overlapping orthogonally oriented lines. Subjects were asked to give introspective reports about where they perceived the orthogonal lines.

They were instructed to guess, if they did not see it. Subjects were first presented with the stimuli in a training round, where stimulus duration was gradually lowered towards threshold.

The experiment was run in three conditions. In the first condition, subjects were instructed to give a response as to where the target had been, after which subjects rated their level of subjective awareness of the stimulus. Subjects pressed one of two keyboard buttons labelled "no" and "yes" as to whether they had seen the stimulus. The response was prompted by the message "Image seen? No or yes?" on the screen.

In the second condition, subjects used PAS ratings. Here, subjects were asked to rate their level of subjective awareness using the PAS scale as mentioned above. The scale was explained to each subject, and a copy was left with them to refer to if needed. After the experiment, subjects were given a short questionnaire to ascertain details, such as how they understood the two different scales and how they felt using them.

At PAS level 1, subjects were at chance level (probability of correct answer was 31%) to almost certain at PAS level 4 (probability of correct answer was 94%). At the same time, PAS fitted better with objective measures such as stimulus duration and correctness than did the dichotomous report. The strongest line of evidence for the validity of the PAS scale as a reflection of conscious perception is however the subjects' own reports that it is the case. Furthermore, subjects often reported that the dichotomous measure was more difficult to use, even though it should seem simpler in terms of numbers in the scale.

One could argue that collapsing PAS levels 1–2 and 3–4, i.e. transforming it into a two-point scale, would essentially lead to a scale identical to the dichotomous scale. Overgaard et al. (2006) calculated for each level of PAS the percentage of cases in which subjects answered 'image seen' when reporting dichotomously (but presented with identical stimulus at equal display time). The results showed that in more than 20% of the cases where subjects reported a PAS score of 1 they responded 'image seen' on the binary scale. Furthermore given a subject reported PAS level 2 the probability that he would answer 'image seen' on the dichotomous scale is 39%. This is one of the more curious aspects of the study, indicating that different processes lie behind reporting in a binary and in a continuous way.

Essentially, the experiment indicates that subliminal perception at least in some cases is a methodological artefact based on the fact that researchers often trust dichotomous subjective reports rather than more elaborate methods to obtain data about conscious states. This, of course, depends on the definition of subliminal perception. As long as subliminal perception is defined as perception in the complete absence of consciousness, the conclusion seems to be correct.

What, then, do the scale points represent more exactly? One strong interpretation of the PAS scale is that consciousness meaningfully can be divided in different "stages", so that, instead of one threshold between conscious and unconscious perception, we would have three thresholds between different subjectively identifiable levels. A different interpretation would simply argue that conscious perception is continuous, and that, for some reason, a scale with four points often seems the most workable in experiments.

Christensen et al. (2006) attempted to shed light on this issue. Using fMRI, they presented subjects with brief stimuli that were rated for experienced clarity individually for each subject. When subjects reported clear conscious experiences, activation was found in a widespread neural network, including partietal, prefrontal, and premotor cortex, supplementary motor areas, insula and thalamus. When subjects reported less clear perceptions, the same areas were less activated and additional unique activations patterns were found. The study supports the idea that there are in fact ontological differences underlying the different PAS scale points. Regarding the exact interpretation of the reports, the study does not provide a solution.

Even more problematic for the concept of subliminal perception in the understanding of Eriksen is a recent study by Overgaard et al. (in prep.) presenting a blindsight patient, TB, who exhibits the subliminal capabilities associated with blindsight using a dichotomous report. However, when the patient was asked to report using PAS, there was a significant correlation between correctness and consciousness in her "blind" field, just as in her "healthy" field.

The Status of Subliminal Perception

So far, we have prepared the case that subliminal perception may not be a real phenomenon at all. Instead, as shown, subliminal perception may be an artifact of (a) the result of objective measures that boil down to other behavioural measures and the a priori assumption of identity between sensitivity and consciousness, and (b) crude subjective measures (e.g. dichotomous or arbitrary 10-point scales) which claim to measure conscious experiences as they are to the subject, but which presumably do not succeed.

One may however raise the contrary argument that we have only shown that subliminal perception does not exist at an "objective identification threshold" but not at an "objective detection threshold". Thus so, since the experiments reviewed above all concern tasks in which the subject is to report the identification of a given stimulus, not just detect if a stimulus was there at all. It has been argued that the identification task is more complex, demands more exposure to the stimulus, and, as a consequence, it is more difficult not to involve subjective experience (Snodgrass et al. 2004a). In a review of the literature on the detection threshold (Snodgrass et al. 2004b), it is shown that there are rather strong statistical arguments that even null sensitivity can be found with relatively low uncertainty.

This interpretation leaves room for subliminal perception as a "fascinating phenomenon", but it still challenges the "tip of the iceberg"-understanding of conscious vs. unconscious perception if all we can do unconsciously is to detect whether something was present or not. However, even this interpretation can be challenged as it is based on a dichotomous understanding of experience. One could interpret the above mentioned PAS findings in such a way that the "weak glimpse"-report actually represents weak experiences at the detection threshold.

At this point, correctness raises significantly above base chance, yet all that is subjectively experienced is a detection that "something was there". Whether this subjective experience correlates with an objective threshold for detection is a still unanswered empirical question. Certainly, however, it must be concluded that until subjective measures with a better fit to the actual experience than what can be achieved with a dichotomous measure are applied, even the existence of a subliminal detection effect is not a settled issue.

References

Baldwin MW (1994) Primed relational schemas as a source of self-evaluative reaction. J Soc Clin Psychol 13:380–403

Balota DA (1983) Automatic semantic activation and episodic memory encoding. J Verbal Lear Verbal Behav 22:88–104

Bargh JA (ed) (2007) Social psychology and the unconscious: the automaticity of higher mental processes. Psychology, Philadelphia, PA

Bargh JA, Chen M, Burrows L (1996) Automaticity of social behaviour: direct effects of trait constructs and stereotype activation on action. J Pers Soc Psychol 71:230–244

Bayne T (2004) Closing the gap? Some questions for neurophenomenology, Phenomenol Cogn Sci 3(4):349–364

Braddock G (2001) Beyond reflection in naturalized phenomenology, J Conscious Stud, 8(11):3–16

Cheesman J, Merikle PM (1984) Priming with and without awareness. Percept Psychophys 36(4):387–395

Cheesman J, Merikle PM (1986) Distinguishing conscious from unconscious perceptual processes. Can J Psychol 40(44):343–367

Christensen MS, Ramsøy TZ, Lund TE, Madsen K (2006) An fMRI study of the neural correlates of graded visual perception. NeuroImage 31(4):1711–1725

Dijksterhuis A (2004) I like myself but I don't know why: enhancing implicit self-esteem by subliminal evaluative conditioning. J Pers Soc Psychol 86:345–355

Dijksterhuis A, Nordgren LF (2006) A theory of unconscious thought. Perspect Psychol Sci 1:95–109

Dixon NF (1971) Subliminal perception: the nature of a controversy. McGraw-Hill, London/New York

Dixon NF (1981) Preconscious processing. Wiley, New York

Draine SC, Greenwald AG (1998) Replicable unconscious semantic priming. J Exp Psychol Gen 127:286–303

Dutton DG, Aron AP (1974) Some evidence for heightened sexual attraction under conditions of high anxiety. J Pers Soc Psychol 30:510–517

Erdelyi MH (1985) Psychoanalysis: Freud's cognitive psychology. Freeman, NewYork

Ericsson KA, Simon HA (1984) Protocol analysis: verbal reports as data. Bradford Books/MIT Press, Cambrdige

Eriksen CW (1956) An experimental analysis of subception. Am J Psychol 69:625–634

Eriksen CW (1960) Discrimination and learning without awareness: a methodological survey and evaluation. Psychol Rev 67:279–300

Goldiamond I (1958) Subliminal perception, subception, unconscious perception: an analysis in terms of psychophysical indicator methodology. Psychol Bull 55(6):373–411

Green DM, Swets JA (1966) Signal detection theory and psychophysics. Wiley, New York

Greenwald AG, Banaji MR (1995) Implicit social cognition: attitudes, self-esteem, and stereotypes. Psychol Rev 102:4–27

Greenwald AG, Draine SC, Abrams RL (1996) Three cognitive markers of unconscious semantic activation. Science 273:1699–1702

Holender D (1986) Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: a survey and appraisal. Behav Brain Sci 9(1):1–66

Holender D, Duscherer K (2004) Unconscious perception: the need for a paradigm shift. Percep Psychophys 66(5):872–881

Holland RW, Hendriks M, Aarts H (2005) Smells like clean spirit: unconscious effects of scent on cognition and behavior. Psychol Sci 16:689–693

Jacoby LL (1991) A process dissociation framework: separating automatic from intentional uses of memory. J Memory Lang 30:513–541

Jacoby LL, Lindsey DS, Toth JP (1992) Unconscious influences revealed: attention, awareness and control. Am Psychol 47:802–809

Kouider S, Dupoux E (2004) Partial awareness creates the "illusion" of subliminal semantic priming. Psychol Sci 15(2):75–81

Leh S, Johansen-Berg H, Ptito A (2006) Unconscious vision: new insights into the neuronal correlate of blindsight using diffusion tractography. Brain 129(7):1822–1832

Macmillan NA (1986) The psychophysics of subliminal perception. Behav Brain Sci 9:38–39

Macmillan NA, Creelman CD (1991) Signal detection theory. Cambridge University Press, Cambridge

Marcel A (1980) Conscious and preconscious recognition of polysemous words: locating the selective effects of prior verbal contexts. In: Nickerson RS (ed) Attention and performance VIII. Erlbaum, Hillsdale

Marcel A (1983) Conscious and unconscious perception: an approach to the relations between phenomenal experience and perceptual processes. Cogn Psychol 15:238–300

Merikle PM (1984) Toward a definition of awareness. Bull Psychon Soc 22:449–450

Merikle PM (1992) Perception without awareness: critical issues. Am Psychol 47:792–795

Merikle PM, Daneman M (1998) Psychological investigations of unconscious perception. J Conscious Stud 5:5–18

Merikle PM, Joordens S, Stolz JA (1995) Measuring the relative magnitude of unconscious influences. Conscious Cogn 4:422–439

Murphy ST, Zajonc RB (1993) Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposure. J Pers Soc Psychol 64(5):723–739

Nisbett RE, Wilson TD (1977) Telling more than we can know: verbal reports on mental processes. Psychol Rev 84:231–259

Overgaard M (2004) On the naturalising of phenomenology. Phenomenol Cogn Sci 3(4): 365–379

Overgaard M, Fehl K, Mouridsen K, Cleeremans A (2008) Seeing without seeing? Degraded conscious vision in a blindsight patient, PLoS ONE, 3(8):1–4

Overgaard M (2006) Introspection in science. Conscious Cogn 15:629–633

Overgaard M, Rote J, Mouridsen K, Ramsøy TZ (2006) Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. Conscious Cogn 15:700–708

Peirce CS, Jastrow J (1885) On small differences in sensation. Mem Nat Acad Sci 3:73–83

Pöppel E, Held R, Frost D (1973) Residual visual function after brain wounds involving the central visual pathways in man. Nature 243:295–296

Ramsøy TZ, Overgaard M (2004) Introspection and subliminal perception. Phenomenol Cogn Sci 3(1):1–23

Raymond JE, Shapiro KL, Arnell KM (1992) Temporary suppression of visual processing in an RSVP task: an attentional blink? J Exp Psychol Hum Percept Perform 18:849–860

Reingold EM (2004) Unconscious perception at the objective detection threshold exists. Percep Psychophys 66(5):888–895

Reingold EM, Merikle PM (1988) Using direct and indirect measures to study perception without awareness. Percep Psychophys 44(6):563–575

Reingold EM, Merikle PM (1990) On the inter-relatedness of theory and measurement in the study of unconscious processes. Mind Lang 5:9–28

Richardson-Klavehn A, Bjork RA (1988) Measures of memory. Annu Rev Psychol 39:475–543

Roy JM, Petitot J, Pachoud B, Varela FJ (1999) Beyond the gap: an introduction to naturalising phenomenology. In: Petitot et al. (eds.) Naturalising phenomenology, Princeton University Press, Cambridge

Schachter S, Singer J (1962) Cognitive, social, and physiological determinants of emotional state. Psychol Rev 69:379–399

Sidis B (1898) The psychology of suggestion. D. Appleton and Company, New York

Snodgrass M, Bernart E, Shevrin H (2004a) Unconscious perception: a model-based approach to method and evidence. Percep Psychophys 66(5):846–867

Snodgrass M, Bernart E, Shevrin H (2004b) Unconscious perception at the objective detection threshold exists. Percep Psychophys 66(5):888–895

Stroh MA, Shaw M, Washburn MF (1908) A study of guessing. Am J Psychol 19:243–245

Timmermans B, Van Duynslaeger M, Vandewaerde N, Van Overwalle F (2008) Your hidden smile or scowl affects me. Subliminally manipulating self-esteem using facial expressions and traits. Manuscript submitted for publication

Varela FJ (1996) Neurophenomenology: A methodological remedy to the hard problem. J Conscious Stud 3(4):330–50

Weiskrantz L, Barbur JL, Sahraie A (1995) Parameters affecting conscious versus unconscious visual discrimination with damage to the visual cortex (V1). Proc Nat Acad Sci USA, 92:6122–6126

IW – "The Man Who Lost His Body"[1]

David McNeill, Liesbet Quaeghebeur, and Susan Duncan

Introduction

Mr. Ian Waterman, sometimes referred to as 'IW', suffered at age 19 a sudden, total deafferentation of his body from the neck down – the near total loss of all the touch, proprioception, and limb spatial position senses that tell you, without looking, where your body is and what it is doing. The loss followed a never-diagnosed fever that is believed to have set off an auto-immune reaction. The immediate behavioral effect was immobility, even though IW's motor system was unaffected and there was no paralysis. The problem was not lack of movement per se but lack of control. Upon awakening after 3 days, IW nightmarishly found that he had no control over what his body did – he was unable to sit up, walk, feed himself or manipulate objects; none of the ordinary actions of everyday life, let alone the complex actions required for his vocation. To imagine what deafferentation is like, try this experiment suggested by Shaun Gallagher: sit down at a table (something IW could not have done at first) and place your hands below the surface; open and close one hand, close the other and extend a finger; put the open hand over the closed hand, and so forth. You know at all times what your hands are doing and where they are but IW would not know any of this – he would know that he had willed his hands to move but, without vision, would have no idea of what they are doing or where they are located.

The IW case is a fascinating study of a person who has lost his body schema (to use Shaun Gallagher's terminology), "his body" in the title of the 1998 BBC *Horizon*

D. McNeill (✉) and S. Duncan
Department of Psychology, S. University Ave., 5848, Chicago, IL, 60637, USA

L. Quaeghebeur
Department of Philosophy, Center for Philosophical Psychology, University of Antwerp, Prinsstraat, 13, 2000, Antwerp, Belgium

program that we quote in our title. The neuronpathy destroyed all sensory neurons at roughly the neck level in proportion to their myelination and conduction speed. The initial medical prognosis was that IW would spend the rest of his days confined to a wheelchair. Not one who takes setbacks lightly, IW commenced a rigorous self-designed and administered program of movement practice with the aim of learning to move again, endlessly performing motions over and over in different combinations, different trajectories, different distances and velocities, until he could, by thinking about the motion and using vision as his guide, plan and execute movements flaw-lessly, so nearly so that observers find nothing unusual about them. Jonathan Cole's book describing IW and his self-administered recovery was called *Pride and a Daily Marathon*, a title that captures in a nutshell the rigor and determination of IW battling the catastrophe that had befallen him (see Cole 1995). After more than 30 years, IW has developed an entirely new way of initiating and controlling movement. He has perfected this style to an astonishing degree. His movements depend on having constant visual contact with the environment, including the surrounding space, objects to be manipulated and any other objects in the immediate vicinity. Every movement is planned in advance, the force and direction calculated intuitively, and the movement monitored as it is taking place. Given all these requirements, it is impressive to see IW move without visible flaw at normal speeds. Although his gait seems somewhat lum-bering (he calls it controlled falling), his arm and hand movements are truly indistin-guishable from normal. However, if vision is denied, IW can no longer control his hands and arms accurately.

Such was the situation in 1997 when, for the first time, the University of Chicago Gesture Lab had a chance to observe IW at first hand. It was through Shaun Gallagher that we had become aware of IW in the first place. Shaun provided a video that had been made some years earlier of IW and another deafferented man during a visit by IW and Jonathan Cole to Pittsburgh. There were ample gesture occurrences on this tape, and so far as we could see IW's gestures appeared com-pletely normal, just as we had observed in the gesturing of unaffected individuals. But of course he had visual contact with his hands at all times. There was accord-ingly the possibility that, like his practical world-related motions, the gestures were planned and monitored. We shall see that indeed some of IW's gestures are created this way. But others are not, and the exceptions are of great interest for the light they shed on how the human brain orchestrates communicative motions in the absence of feedback.

Shaun also put us in touch with Jonathan Cole and from him we received a copy of *Pride*, which had already been published in England. These preliminary contacts began as early as 1992. We had begun to ponder what IW would do gesturally if we could remove visual contact with his arms and hands. The BBC *Horizon* program provided our opportunity, flying IW, Jonathan and Shaun to the University of Chicago for experiments to our devising while they filmed us. Shaun Gallagher's book (2005), *How the body shapes the mind*, has a full chapter devoted to these experiments. We shall cover much the same ground but with a more psycholinguistic and gesture-study slant; descriptions fully complementary to Gallagher's.

The Study of Gesture and Its Implications

Below we describe these experiments and others performed on another visit, but first we explain briefly the kinds of gestures we focus on, how we study them, and what they reveal of a specific mode of cognition during speech.

The Gesture Continuum

The word 'gesture' covers a range of communicative events. The term is nonetheless convenient and we shall retain it for this chapter, but first we draw some crucial distinctions. The gestures of concern to us are an integral component of language, not a substitute, accompaniment or ornament. Such gestures are synchronous and co-expressive with speech, not redundant, and not signs, salutes, or so-called emblems (see below). They are frequent – about 90% of spoken utterances in descriptive discourse are accompanied by them (Nobe 2000). They occur in similar form across many cultures (we have observed speakers from more than 20, including 'high-gesture' cultures, such as Naples). The gestures so described were termed 'gesticulations' by Kendon (1988); other gestures in his terminology were 'language-like' and 'pantomime' – all contrasted to 'signs'. Arranged on a continuum, they can be organized as follows (McNeill 1992):

Spontaneous Gesticulation → Language-like → Pantomime → Emblems → Signs

The differences along The Gesture Continuum map onto three dimensions – how necessary speech is to the gesture; how language-like is the gesture; and how conventionalized is its form. These three perhaps can be reduced to an unnamed deeper dimension. Nonetheless, it is useful to see how points on the Continuum differ on the three. So as one goes from gesticulation to sign the relationship of gesture to speech changes.

• The obligatory presence of speech declines.
• Language-like properties increase.
• Socially regulated conventional signs replace self-generated form-meaning pairs.

Language-like gestures have a different timing relationship with speech from gesticulations. For example in, "he goes [-]," a gesture synchronizes with a momentary pause in speech; a vacant grammatical slot. Here gesture substitutes for speech. An emblem is a culturally established morpheme (or semi-morpheme, because it does not usually have combinatoric, 'syntagmatic' values), such as the "OK" sign and others. Emblems can occur with or without speech. Pantomime is gesture without speech, often in sequences and usually comprised of simulated actions. What distinguishes pantomime from gesticulation is that the latter, but not the former, is integrated with speech. Pantomime, if it relates to speaking at all, does so as a 'gap

filler' (to use a phrase by Susan Duncan, pers. comm.). Speech-gesticulation combinations are cognitive constructions, and occur where speech and gesture are co-expressive of the same idea.[2] Sign languages are full, socially constituted, non-spoken languages.

Even though 'gesticulation' (hereafter, 'gesture') is only one point on the Continuum, it dominates gesture output in storytelling, living space descriptions, academic discourse (including prepared lectures) and conversations. Such gestures synchronize with speech at points where they and speech embody shared underlying meanings in discourse, possess "communicative dynamism" (Firbas 1971), and are points of maximal discursive force (McNeill and Duncan 2000). Commonly 99% if not all gestures in such contexts count as 'gesticulation'. An example from a student participant in one of our earliest experiments is shown in Fig. 1.[3]

Gestures and Speech – Two Simultaneous Modes of Semiosis

Figure 1 illustrates synchronous co-expressive speech and a gesture recorded during a narration. The speaker had just watched a cartoon and was recounting it to a listener from memory. We explained that the task was storytelling and did not mention gesture (the same method was used with IW). The speaker was describing an event in which one character (Sylvester) attempted to reach another character

Fig. 1 Gesture combining entity, upward movement and interiority in one symbol

---

[2]Movement by itself offers no clue to whether a gesture is 'gesticulation' or 'pantomime'; what matters is whether the two modes of semiosis, linguistic form and gesture, simultaneously co-express one idea unit.

[3]Computer art from video by Fey Parrill, Ph.D. Except for Fig. 9, all illustrations are from McNeill (2005), *Gesture and Thought* (University of Chicago Press), and are used with permission.

(Tweety) by climbing up the inside of a drainpipe; a pipe conveniently topping out next to a window where Tweety was perched. The speaker said, "and he goes up thróugh the pipe this time." Co-expressively with "up" her hand rose and with "thróugh" her fingers spread outward to create an interior space. The upward movement and the opening of the hand were simultaneous and both synchronized precisely with "up thróugh," the linguistic package that carried the related meanings. The prosodic emphasis on "thróugh," highlighting interiority, is matched by the added complexity of the gesture, the spreading and upturning of the fingers. What we mean by co-expressivity here is this joint highlighting of the ideas of rising and interiority, plus their joint contribution to communicative dynamism.[4]

However, also note the differences between the two types of semiosis. Speech componentializes the event: a directed path ("up") plus the idea of interiority ("thróugh"). This analytic segregation further requires that direction and interiority be concatenated, to obtain the composite meaning of the whole. In contrast, gesture is a synthesis. The whole emerges as one symbol. The semantic elements in speech are simultaneously aspects of this imagery whole. No concatenation is required. Meaning determination moves from whole to parts, not from parts to whole. The effect is a uniquely gestural way of packaging meaning – something like "rising hollowness." Thus, speech and gesture, co-expressive but non-redundant, represent one event (climbing up inside the pipe) in two forms: analytic/combinatoric and global/synthetic – at the same instant.

This kind of gesticulation is also our focus in the IW case. IW is unquestionably capable of combinations of unlike semiotic modes of these kinds in packaging meanings. It is important, however, to register a distinction within the gesticulation type introduced by IW himself. Some of his gestures, he says, are constructed: planned in advance, launched at will, and controlled in timing and motion throughout – carried out, in other words, exactly as he carries out his practical, world-related movements. His second type he calls 'throw-aways' – "ones that just happen. Sometimes I'll be aware of them because there may be something around me … but most are just thrown away." 'Throw-aways' are not explicitly planned and monitored, and precisely for this reason are of great interest.

The Binding of Speech and Gesture

A final point is the binding of gestures and speech when they participate in the formation of cognitive units, a binding so strong that efforts to separate them fail – either speech and gesture remain together or they are jointly interfered with; in either case the speech-gesture bond is unbroken. We expect the same to hold with IW's 'throw-away' gestures (his 'constructeds,' arising from deliberate planning, generally do not show the same binding with speech). The following are experimental examples of tight binding gleaned independently of IW from the gesture literature:

---

[4]More extensive accounts are in McNeill (1992) and McNeill (2005).

- Delayed auditory feedback – the experience of hearing your own voice played back after a short delay – produces major speech disturbances but does not interrupt speech-gesture synchrony (McNeill 1992).
- Stuttering and gesture are incompatible. The onset of a gesture inoculates against stuttering and, conversely, the onset of stuttering during a gesture interrupts it instantly (Mayberry and Jaques 2000).
- People blind from birth, who have never seen gestures and have no benefit from experiencing them in others, gesture and do so even to other blind people whom they know to be blind (Iverson and Goldin-Meadow 1997).
- Memory loss interrupts speech and gesture jointly; it is not that gesture is a 'gap-filler' when memory fails (McNeill 2005).
- Conversely, gestures protect memory from interference (Goldin-Meadow et al. 2001).

The speech-gesture units in these settings are held together by the requirements of idea unit formation: thought in speech takes place simultaneously in imagery and linguistic form; to think while speaking is to be active in both these modes at once. Speech and gesture are thus yoked, because both are essential to this distinctive form of cognition. For a recent statement of a 'growth point' hypothesis that explains this double essence of thinking while speaking, see McNeill et al. (2008). We return to the growth point at the end of this chapter.

IW's gestures

The BBC brought IW, Jonathan Cole and Shaun Gallagher to the University of Chicago for filming in July 1997. We wanted to record IW under a variety of conditions, both with and without vision. IW cannot be simply blindfolded, since he would be unable to orient himself and be at risk of falling over. We devised a tray-like blind, pictured in Fig. 2, that could be pulled down in front of him, blocking vision of his hands, while allowing him space to move and preserving his visual contact with his surroundings.[5] IW was videotaped retelling the above-described animated cartoon. He also was recorded under the blind in casual conversation with Jonathan Cole. In 1997, we did not appreciate the importance of testing IW's instrumental actions without vision but we had an opportunity to test his performance on this kind of task in April 2002, when IW and Cole came back for a second visit to the University of Chicago.[6]

---

[5] Nobuhiro Furuyama suggested the blind experiment. The blind was designed and built by David Klein.

[6] The second round of experiments was supported by a grant from the Wellcome Trust to Jonathan Cole and by funds from Ian Waterman.

Fig. 2 IW seated at the blind designed for gesture experiments

Significant Variables in Assessing IW's Gesture Performance

To have a systematic approach to IW's gestures, we pay specific attention to the following variables:

Timing: synchronization with co-expressive speech
Morphokinesis: the shape of the gesture in terms of hand forms and use of space
Topokinesis: the location of hands relative to each other in space including but not limited to the approach of one hand by the other
Character viewpoint (CVPT): the perspective of the character being described a gesture from the CVPT is close to mimicry
Observer viewpoint (OVPT): the perspective of the narrator or an observer

With vision, IW's gestures display all the above features (over a sample of gestures). Without vision, they show some but not all features: exact timing with speech, morphokinetic accuracy, and OVPT. Topokinetic accuracy and CVPT, however, become rare. The loss or reduction of these two particular features implies that his gestures without vision depart from the pathway of world-related action control (regarding CVPT as mimicry or action simulation). The preservation of speech-gesture synchrony implies that the system that remains is integrated with speech. The ensemble of preserved and lost features suggests a dedicated thought-language-hand link. This link will be discussed in more detail later in the chapter.

IW's Gestures With and Without Vision (1997)

IW's gestures with vision are similar to those produced by normal speakers, although they are fewer in number and tend to be isolated, performed one by one, in keeping with his self-conscious constructed-gestures strategy. Figure 3 shows a

Fig. 3 IW iconic gesture with vision
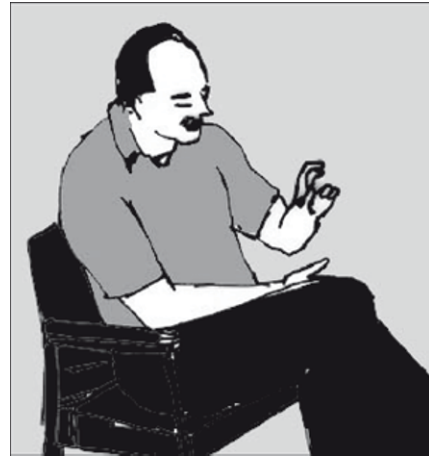
narrative gesture made with vision. IW was describing Sylvester after he had swallowed a bowling ball that Tweety had dropped inside the pipe. Both morphokinesis and topokinesis are indistinguishable from normal. His hand appears to bracket a small figure in the central gesture space and move it downward, wobbling right and left slightly as it went down. The motion is co-expressive with the synchronous speech: [**//he // wiggles his way down**] (bold face indicates speech accompanying gesture). The only clue that control is other than normal is that IW looks at his hand during the gesture. The viewpoint in this case is that of an observer; elsewhere, in his full description of the bowling ball episode, the character viewpoint also occurs (OVPT and CVPT refer to the accompanying gestures, not the spoken forms).

OVPT: "tiny little bird" – left hand appears to outline bird (cf. Fig. 3)
CVPT: "bowling ball" – both hands appear to thrust down on ball
OVPT: "wiggles his way down" – left hand again outlines bird, wiggles
CVPT: "places it" – left hand appears to push down ball
OVPT: "gets a strike" – hands move laterally from center space

Figure 4 illustrates a narrative gesture without vision, a coordinated two-handed tableau, in which the left hand is Sylvester and the right hand is a trolley pursuing him. IW was saying, "[and the $_a$tram $_b$**caught** *him up*]" (a, b referring to the first and second panels of the illustration). His right hand moved to the left in exact synchrony with the co-expressive "caught". Moreover, a poststroke hold extended the stroke image through "him" and "up" (underlining) and thus maintained full synchrony of the meaningful configuration in the stroke with still unfolding co-expressive speech. It is important to recall that this synchrony and co-expressivity were achieved without proprioceptive or spatial feedback. We thus see in IW, without any feedback, the double semiosis of synchronous gesture and speech.

Figure 4 demonstrates another similarity of IW's 'throw-aways' to normal gestures. The gesture is complex, it uses two hands doing different things in relation to each

other, the whole imagery depicting a situation in which the entities identified in speech are changing their relationships in time and space. Such complexity contributes to communicative dynamism; that is the case with Fig. 4 – the event is the denouement of a buildup and the main discursive point.

Topokinetic Versus Morphokinetic Accuracy

The gesture in Fig. 4 was accurate morphokinetically but not topokinetically; as the right hand approached the left, the right and left hands did not line up. Figure 5 illustrates another case of topokinetic approximation. IW was describing "a square plank of wood" and sketched a square in the gesture space. The illustration captures the misalignment of his hands as he completed the top of the square and was about to move both hands downward for its sides.
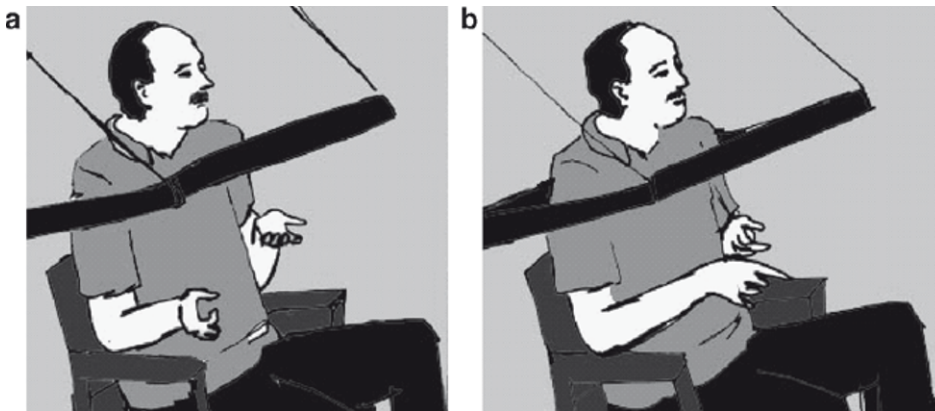
Fig. 4 (**a**, **b**) IW coordinated two-handed iconic gesture without vision

Fig. 5 Lack of topokinetic accuracy without vision
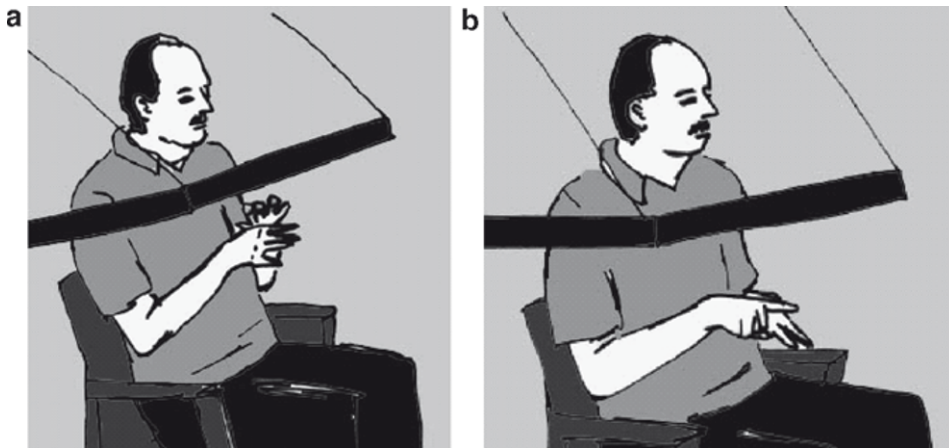
Fig. 6 (a, b) IW's misalignment as he outlines a triangle and circle without vision

Fig. 7 Accurate completion of triangle by subject with intact proprioception and spatial sense without vision

We also asked IW to sketch simple geometric shapes in the air without vision. Morphokinetically, a triangle and a circle were readily created but topokinetically there was always some disparity (Fig. 6a, b show the end positions of a triangle and circle, respectively). For comparison, we also asked undergraduate students at the University of Chicago to sketch geometric figures without vision. Figure 7 is the end point of one such sketch of a triangle. Positioning is exact to the millimeter.

Instrumental Actions

Similarly, instrumental actions denied vision are difficult for IW. Such actions require topokinetic accuracy. Figure 8 shows two steps in IW's attempt to remove the cap from a thermos bottle. The first is immediately after Jonathan Cole has placed the thermos in IW's right hand and placed his left hand on the cap (IW is strongly left handed); the second is a second later, when IW has begun to twist the cap off. As can be seen, his left hand has fallen off and is turning in midair. Similar disconnects without vision occurred during other instrumental actions (threading a
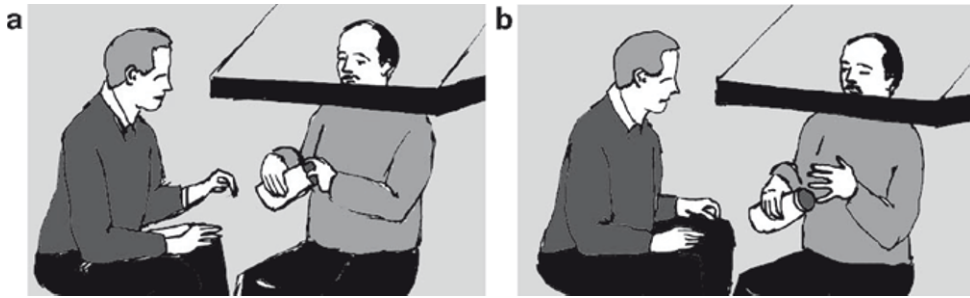
Fig. 8 (**a**, **b**) IW attempts to perform an instrumental action (removing cap from a thermos)

cloth through a ring, hitting a toy xylophone, etc. – this last being of interest since IW could have made use of acoustic feedback or its absence to know when his hand had drifted off target, but still he could not perform the action).

Significance of the IW Results So Far

The IW case shows that, without vision, gestures continue to occur with accuracy up to the morphokinetic level, and possess the tight binding at points of co-expression with speech that characterizes unaffected gestures – all this without feedback of any kind. An important hypothesis is that a dedicated thought-language-hand brain link underlies combinations of semiotically unlike meaning packages that can be partially dissociated from the brain circuits involved in world-related actions. IW's use of space is especially informative. Although he has no exact sense of where his hands are, he can align them morphokinetically to create a 'triangle', because triangularity affords a direct mapping of a concept into space. Likewise, the meaning of "catching up to" is sufficient to guide the hands into a morphokinetic embodiment of this idea, without an intervening action, real or simulated (cf. discussion in Gallagher 2005).

The morphokinetic/topokinetic distinction also explains the near disappearance of CVPT without vision. Gestures like 'holding it' and 'places it', with CVPT, resemble Tweety's instrumental actions of holding the bowling ball and placing it. These CVPT gestures have meanings as simulated actions of a kind that require the level of control that, for IW, only visual guidance provides. Hence they become difficult when vision is absent.

IW Can Control Speech and Gesture in Tandem (1997)

A striking demonstration of the thought-language-hand link is that IW, without vision, can modulate the speed at which he presents meanings in both speech and

gesture, and do this *in tandem*. As his speech slows, his gesture slows, too, and to the same extent, so that speech-gesture synchrony is exactly preserved. If what he is forming are cognitive units comprised of co-expressive speech and gesture imagery in synchrony, this joint modulation of speed is explicable. He does it based on his sense (which is available to him) of how long the joint imagery-linguistic cognitive unit remains 'alive'; peripheral sensory feedback need not bepart of it. During a conversation with Jonathan Cole while still under the blind, IW reduced his speech rate at one point by about one-half (paralinguistic slowing), and speech and gesture remained in synchrony:

Normal: "and I'm startin' t'use m'hands now"

Slow: "because I'm startin' t'get into trying to explain things"

The gestures are of a familiar metaphoric type in which a process is depicted as a rotation in space (possibly derived from ancient mechanisms, perhaps mill-wheels or clockworks: metaphoric gestures often freeze-dry images that exist now only in this gesture form; cf. McNeill 1992 for other examples). IW executes the metaphor twice; first at normal speed, then at slow speed. The crucial observation is that the hand rotations are locked to the same landmarks in speech despite the different speeds. IW's hands rotate in phase at normal speed, opposite phase at slow speed. Nonetheless, if we look at where the hands orbit inward and outward we find that rotations at both speeds coincide with the same lexical words, where they exist, and with the same stress peaks throughout. Figure 9 shows the maximum inward and outward hand motions and the coincident speech.[7] Brackets indicate where linguistic content was identical at the two rates.

This agreement across speeds shows that whatever controlled the slowdown, it was exactly the same for speech and gesture. Bennett Bertenthal (pers. comm.) points out a possible mechanism for this tandem reduction. Speech and gesture, slowing together, could reflect the operation of a pacesetter in the brain that survived IW's deafferentation; for example, the hand moves outward with a peak, an association that could be maintained over a range of speeds. The rotating hands were as noted metaphors for the idea of a process. The pacesetter accordingly could be activated by the thought-language-hand link and co-opted by a significance other than the action of rotation itself. This metaphoric significance is consistent with the timing, since the hands rotated only while IW was saying "I'm starting to …" and there was actually a cessation of gesture between the first (normal speed) and second (reduced speed) rotations as he said "and that's because", indicating that the rotation and any phonetic linkages it claimed were specifically organized around presenting the idea of a process as a rotation in space.

---

[7] The presentation of speech and gesture events in this figure, by adding an extra panel and selecting the limits of rotation in each panel, improves accuracy without changing the analysis from that in McNeill (2005).

Normal Speed (bracketed material=0.56 sec., 5 syllables) "and [I'm startin' t'] use m'hands now"		Slow Speed (bracketed material=0.76 sec., 5 syllables) "-cuz [I'm startin' t'] get into"
	and I'm ⇐ 'cuz I'm ⇒	
	startin' ⇐ startin' ⇒	
	t' use m- ⇐ t' get in- ⇒ (at right, hands rotating out of phase, left hand rotates max in, right hand max out; corresponds to both hands max in at left, with hands rotating in phase)	

Fig. 9 IW changes rate of speech and gesture in tandem, maintaining synchrony. Note that motion of hands outward and inward occurs at same speech points
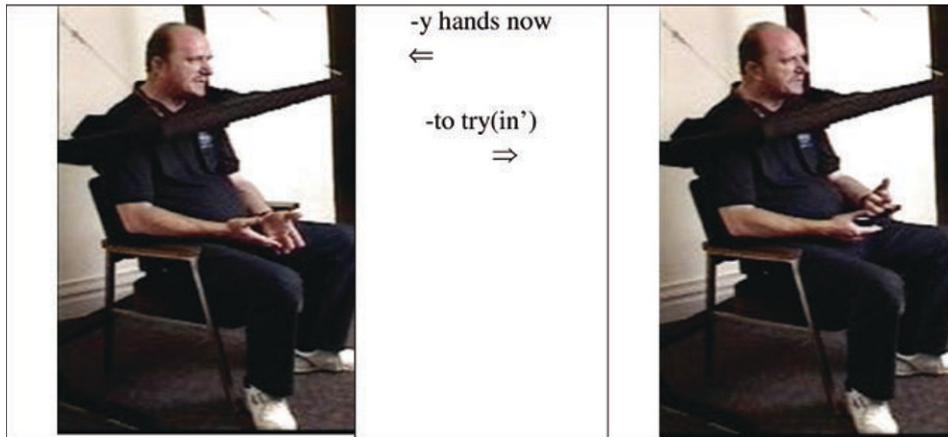
Fig. 9 (Continued)

Summary of IW's Gestures Without Vision

The following points summarize what we have seen of IW's gestures in the absence of visual, proprioceptive or spatial position feedback:

- Gestures have diminished CVPT.
- Gestures preserve morphokinetic accuracy and lose topokinetic accuracy.
- Gestures are co-expressive and synchronized with speech.

Phantom Limb Gestures

V.S. Ramachandran and S. Blakeslee in *Phantoms in the Brain* (1998) describe Mirabelle, a young woman born without arms. Yet she experiences phantom arms and performs 'gestures' with them – nonmoving gestures, but imagery in actional-visual form.

> Dr: "How do you know that you have phantom limbs?" M: "Well, because as I'm talking to you, they are gesticulating. They point to objects when I point to things."

"When I walk, doctor, my phantom arms don't swing like normal arms, like your arms. They stay frozen on the side like this" (her stumps hanging straight down). "But when I talk, my phantoms gesticulate. In fact, they're moving now as I speak." Ramachandran and Blakeslee (1998, 41)

Mirabelle's case points to a similar conclusion as IW's – dissociation of gesture from practical actions. In Mirabelle's case, moreover, intentions create the sensation of gestures when no motion is possible. Presumably, again, the same thought-language-hand link is responsible.

Overall Significance of the IW Case

The IW case suggests that control of the hands and the relevant motorneurons is possible directly from the thought-linguistic system. Without vision, IW's dissociation of gesture, which remains intact, and instrumental action, which is impaired, implies that the "know-how" of gesture is not the same as the "know-how" of instrumental movement (using Shaun Gallagher's terms). In terms of brain function, this implies that producing a gesture cannot be accounted for entirely with the circuits for instrumental actions; at some point the gesture enters a circuit of its own and there is tied to speech. A likely locus of this dedicated thought-language-hand link in the brain areas 44 and 45: Broca's area. The earlier mentioned paper by McNeill et al. (2008) presents a theory of how this link could have been evolutionarily selected in this brain area (called the 'Mead's Loop' model in the paper).

Conclusion: Growth Points, Material Carriers, and Inhabitance

To conclude this chapter we describe the *growth point* (GP) hypothesis mentioned briefly earlier; the concept of a *material carrier* from Rieber and Carton (1987); relate these to the concept of *inhabitance* from Merleau-Ponty (1962) while elaborating somewhat on the phenomenology of gesture; and explain the interconnections among all three concepts as they apply to the IW case.

The Growth Point

It is beyond doubt that IW, at least in his 'throw-aways', is creating what we term growth points. GPs organize speech and thought. A GP is an irreducible, 'minimal unit'[8] of imagery-language code combination. It is the smallest packet of an idea unit encompassing the *unlike* semiotic modes of imagery and linguistic encoding that we observe when speech and gesture coincide at points of co-expressiveness. A GP is empirically recoverable, inferred from speech-gesture synchrony and co-expressiveness. It is inferred (not 'operationally defined') from (a) gesture form, (b) coincident linguistic segment(s), (c) co-expression of the same idea unit, and (d) what Vygotsky (1987, p. 243) termed a 'psychological predicate' – the point of newsworthy content that is being differentiated from the immediate context of speaking (of which, more below).

---

[8]The concept of a 'minimal unit' with the property of being a whole is from Vygotsky (1987, pp. 4–5).

The temporal and semantic synchronies represented in Fig. 1 and shown dramatically by IW when he reduced speed in speech and gesture in tandem, imply a GP in which imagery and linguistic information are jointly present, so that one does not occur without the other. In Fig. 1 we infer the simultaneous presence of the idea of ascent inside the pipe in two unlike semiotic modes. Even when the information ('semantic content') in speech and gesture is similar, it is formed according to contrasting semiotic modes.

The GP is so named because it is a distillation of a growth process – an ontogenetic-like process but vastly sped up and made functional in online thinking-for-speaking. According to this framework, it is the initial unit of thinking-for-speaking (Slobin 1987) out of which a dynamic process of utterance-level and discourse-level organization emerges. Imagery and spoken form are mutually influencing. It is not that imagery is the input to spoken form or spoken form is the input to imagery. The GP is fundamentally both.

The existence of simultaneous unlike modes creates instability; an idea in two contending forms at once. This instability nonetheless is an essential part of the GP and its role in speaking and thought – it drives thinking-for-speaking to seek resolution (McNeill and Duncan 2000).[9] For modern humans, stability comes from 'unpacking' the growth point into grammatical structures (or viable approximations thereto) with usually further meaning generation actualized. A surface linguistic form emerges that cradles the GP in stable and compatible form. This role of grammar – unpacking and supplying 'stop-orders' for the changes initiated by imagery-linguistic code instability – is an important clue for how speech in discourse is produced (see McNeill 2005 for detailed discussion).[10] In Fig. 1, the locution "up thróugh" is analytic: upness and interiority are separated. The words also have syntagmatic values acquired from combinations within and beyond the phrase. The gestural image embodies the same information in the form of 'Sylvester as rising hollowness' but without analysis or combinatoric value. Unpacking resolves the tension by placing both components, linguistic and gestural, into a finished syntactic package that does not violate the image, realizes the syntagmatic potential of the linguistic side, and includes the production of further content ("he goes up through it this time," includes the metanarrative indexical, "this time," that relates the event to a previous one).

A final point is that we can fully understand what motivates any image-speech combination only with reference to how a GP relates to its context of occurrence. The GP-to-context relationship is mutually constitutive. The GP is a 'psychological predicate' – the point of differentiation from this context. The speaker so represents

---

[9]The reasons why semiotic opposition creates instability and initiates change include:

(a) Conflict (between semiotic modes: analog imagery/analytic categorical), and

(b) Resolution (through change: fueling thinking-for-speaking, seeking stability)

Simultaneous semiotic modes comprise an inherently dynamic psycholinguistic model.

[10]When gesture and speech synchronize, as in Figs. 1 and 4a, b, the two modes are in direct contact. If there is less than perfect synchrony, the 'double essence' of the same meaning in unlike semiotic modes can still stimulate unpacking. The ultimate criterion is whether an idea is embodied in two modes (with or without different aspects of the idea) that creates instability.

the context that this differentiation becomes possible. A robust phenomenon concerning gesture is that the form and timing of gestures select just those features that differentiate the psychological predicate in a context that is at least partly the speaker's own creation (see McNeill 2005, pp. 108–112).

We observe all these hallmarks of GPs, including this correlation, in IW's speech and gesture. The "caught him up" gesture, for example, was a psychological predicate that embodied newsworthy content in a context from the preceding narrative discourse of Sylvester on overhead wires running to escape a pursuing trolley. The gesture depicted the pursuit and overtaking by the trolley and was exactly synchronous with the linguistic segments "caught him up." The GP as inferred is this combination of semiotic modes for the idea of Sylvester being overtaken. The unpacking into "and the tram caught him up" settles it into a stable syntactic package (the next element in IW's tale describes how he was then shocked – another GP ensued with its instability followed by stability through unpacking).

Material Carriers

We get a deeper understanding of such an imagery-language dialectic by introducing the concept of a 'material carrier'. The concept clarifies reasons why IW, despite his careful attention to movement up to and including the construction of gestures, yet performs, without meaning to, unattended 'throw-aways'. A material carrier – the phrase was used by Vygotsky (1987)[11]– is the embodiment of meaning in a concrete enactment or material experience.[12] A material carrier appears to enhance the symbolization's representational power. The concept implies that the gesture, *the actual motion of the gesture itself*, is a dimension of meaning.

---

[11]Pointed out by Elena Levy. The quote (recovered thanks to Tae Kunisawa) is "That which is specific to this particular form of sound has remained unexplored. As a consequence, this research has not been able to explain why sound possessing certain physical and mental characteristics is present in human speech or how it functions as a component of speech. In a similar manner, the study of meaning has been defined as the study of the concept, of the concept existing and developing in complete isolation from its material carrier. To a large extent, the failure of classic semantics and phonetics has been a direct result of this tendency to divorce meaning from sound, of this decomposition of the word into its separate elements." (Rieber and Carton 1987, p. 46).

[12]As suggested by the semantic satiation phenomenon (Severance and Washburn 1907, recovered thanks to Fey Parrill): staring at TREE, say, soon disrupts the word. It ceases to be a meaningful symbol and its actual material form seems to change perceptually. A 'satiation effect' for gestures would be equally interesting to document. Trying such an experiment, David McNeill deliberately repeated a gesture that seems typical of him (a gesture for the concept of a growth point, no less; see Parrill 2007), and almost immediately experienced a shift from significant symbol to mere hand rotation for the movement. If vulnerability to semantic satiation indicates the strength of the material carrier in a symbol, this gesture is strong indeed.

The appearance of enhancement is possible if the gesture *is* the very image; not an 'expression' or 'representation' of it, but *is* it. From this viewpoint, a gesture is an image in its most developed – that is, most materially, naturally embodied – form. The absence of a gesture is the converse, an image in its least material form.[13] We describe here a theoretical model of how materialization has this effect on representational power, and when gestures do and do not occur with speech (cf. Goldin-Meadow 2003). A striking illustration of the material carrier is what Cornelia Müller (2008) Snodgrass et al. 2004a terms the 'waking' of 'sleeping metaphors' – new life given to inactive metaphors, in which the gesture brings back to awareness the metaphor's original source. This enhancement shows the gesture as a material carrier. Müller gives an example of a German metaphor ("gefunkt", 'sparked', the equivalent to English 'clicked', for suddenly falling in love). The expression is usually hackneyed and not apprehended as a metaphor. However, it can be awakened by a gesture. A speaker, describing her first love, said "between us somehow it sparked ['clicked']" (Müller's translation). As she said "between us" her hand rose upward next to her face in a ring shape but with an unusual orientation – the fingers pointing at her own face; then, as she uttered the metaphor itself, "gefunkt", her hand abruptly turned outward – her gesture materializing the 'dead' metaphor as a sudden event, an electrical spark.[14]

IW shows the reality of materialization in yet another form. At one point in the 2002 experiment Jonathan Cole demonstrated, as IW watched, an object-directed transitive action (removing the cap from the thermos); IW then imitated the action. While he could not perform the action himself without vision (Fig. 8), we were interested in seeing if he could imitate it under conditions where topokinetic accuracy was not a factor, and indeed he could. But what was unexpected is that IW spontaneously spoke as he imitated the cap removal (he described his movements as he performed them). It was a fully spontaneous and unanticipated performance, not something we suggested, even though, of course, this spontaneous sprouting of speech is what the GP hypothesis implies – two forms of materialization co-occurring.

The inverse experiment happened equally accidentally in a separate study of IW by Bennett Bertenthal (personal communication). Here, too, imitation was the task (he was shown a video, without sound, of other people's gestures and asked to imitate them). As before, IW spontaneously began to speak. The experimental assistant asked him to not speak, as that was not part of the experimental protocol. IW complied and – the important observation for material carrier purposes – his imitations immediately *simplified and shrank dramatically in size*.

---

[13] The material carrier concept thus helps explain why sometimes there is no gesture. When no gesture occurs, we witness the *lowest level of materialization*.

[14] Müller views the metaphor dynamically, as a *process* by which the speaker and her listener generate metaphoricity in the context of the speech event; clearly a conception germane to the position of this book. The activation of the metaphor, and the semiotic impact of the sparking image, is a variable, dependent upon the speaker's thought processes and the context of speaking. The gesture, as a material carrier, is an active component of this process.

Whereas, with speech, they had been large, complex and executed in the space in front of his body (he was not under the blind), without it they were simple, miniaturized and confined to the space at his lap. This was so even though imitation of other people's gestures was his target and he had vision of his hands.

These effects are impressive indications that two materializations, speech and gesture, co-occur, support and feed one another and that when one goes awry or missing the other tends to follow.

Phenomenology and the Scientific Study of Gesture

The entire conception of speech and gesture is moved to a new level when we draw on the work of Maurice Merleau-Ponty for insight into the thought-language-hand link and the temporal alignment of speech, gesture and significance into GPs. First, however, we have to elucidate the situation of present-day gesture studies with respect to the notoriously difficult relationship between phenomenology and (cognitive) science. Merleau-Ponty for one makes a specific distinction between his philosophy of embodiment and the empirical-scientific approach to the role of the body in language use and cognition in general.

Empirical conceptions tend to focus on the body-as-object and describe embodied language use in terms of its objective features, such as the speech sounds uttered, the specific gestures which were made or found patterns of neurological activity. In a two or more step process, the speaker – or rather her cognitive system – embodies some pre-existing meaning (a 'thought') through the realization of complex combinations of different kinds of material carriers (such as the verbal, the manual, the facial and the postural modality), and thus linguistic meaning is 'externalized'. In this approach, the body in language use functions as a machine that can talk, a machine that can 'translate' a private and disclosed thought into the conventionalized medium of material carriers. This kind of mechanistic communicative theory naturally follows from a framework that describes the linguistic event solely from a third-person point of view. The empirical scientist takes a neutral stance vis-à-vis the object of her investigation, i.e. people involved in a conversation *over there*, and she relies on inference in order to discover what goes on when people talk.

Merleau-Ponty on the other hand, being a member of the phenomenological tradition, stresses the importance of acknowledging first-person experience (a description of the body-as-subject) whenever the issue of meaning (perceptive as well as linguistic) is concerned. From this perspective, we do not have the sensation that the speaker's expressive body *mediates* between her thoughts and the listener's cognitive capacities, but on the contrary, we experience that we have a *direct access* to each other's intentions. Embodied meaning makes immediate sense from the perspective of the speaker and the listener. In fact, in this account, meaning coming into existence, its bodily expression and, in a sense, even meaning reception, are *one and the same thing and happen in one and the same instance*. In contrast with empirical conceptions, here the speaking subject (a 'first person') does not *provide* her

thoughts with a material carrier, nor does the listening subject *infer* meaning from her objective perception of someone's expressive bodily movements. Phenomenological embodiment of linguistic meaning is fundamental, it is an a priori fact: the mental (a 'thought', or 'intentional content') and the physiological (its material carrier) are co-emergent (in Heidegger's terms they are 'equiprimordial'). The emergence of meaning and its bodily expression therefore can be said to constitute two aspects of one and the same phenomenon, viz. the speaker's bodily existence in a world which she makes part of and hence to which she is fully attuned. In the next paragraphs we will discuss some implications of this phenomenological framework for any theoretical account of gesture in general and for the case of IW in specific.

With regard to these theoretical antipoles, the empirical-scientific third-person perspective and the phenomenological first-person perspective, where should we locate an approach to gesture that propounds a thought-language-hand link to account for the synchronization of what Duncan (2006) has called the 'three rhythmic pulses': speech, gesture and significance? Lived experience, despite its importance for the understanding of multimodal co-expressivity, by definition cannot be exhaustively described from an objective point of view, but taking a third-person stance is exactly one of the defining traits of the scientific métier – and also that of a science of gesture. Language use necessarily precedes doing linguistics and the unmediated way in which the speaker and her listener grasp the integrated communicative event can, after the fact, never be paralleled by listing the objective features of that event. Only the linguistic subject, because of her non-scientific but actively engaged stance and her ability to grasp a situation *all at once*, can understand the true nature of meaning. The thought-language-hand link, with its power of co-expressiveness, is a suitable way of scientifically approaching this '*all at once-ness*', which exclusively objective descriptions of the role of the body in language use traditionally have difficulties to grasp.[15] Because the thought-language-hand link by definition both distinguishes and equates the three pulses, thus fulfilling both scientific and phenomenological aspirations, it enables us to operationalize the abstract, philosophical concept of the body-as-subject by being capable of inspiring empirical, experimental research.

For a first investigation into the philosophical significance of gesture, we may turn to Merleau-Ponty's *Phenomenology of Perception* (1962) to give us insight into the duality of gesture and language and the ontological status of the GP – its multifaceted cognitive or perceptive way of being. Gesture, the instantaneous, global, nonconventional component, is "not an external accompaniment" of speech, which is the sequential, analytic, combinatoric component; it is not a "representation" of meaning, but instead meaning "inhabits" it:

---

[15] Most empirical-scientific conceptions – at least implicitly – infer that because (and after the fact of speaking itself) a communicative event can be divided up into different aspects by the linguistic scientist, a cognitive system necessarily also must process these aspects one by one (and therefore consecutively) before finding ways of integrating them into a coherent interpretation.

"The link between the word and its living meaning is not an external accompaniment to intellectual processes, the meaning inhabits the word, and language 'is not an external accompaniment to intellectual processes'.[16] We are therefore led to recognize a gestural or existential significance to speech … Language certainly has inner content, but this is not self-subsistent and self-conscious thought. What then does language express, if it does not express thoughts? It presents or rather it *is* the subject's taking up of a position in the world of his meanings." (p. 193; emphasis in the original)[17]

The GP is a mechanism geared to this "existential content" of speech – this "taking up a position in the world". Gesture, as part of the GP, is inhabited by the same "living meaning" that inhabits the word (and beyond, the discourse). A deeper answer to the query, therefore – when we see a gesture, what are we seeing? – is that we see part of the speaker's current cognitive being, her very mental existence, at the moment it occurs. This applies equally to all speakers, IW included. By performing the gesture, a core idea is brought into concrete existence and becomes part of the speaker's own cognitive bodily existence at that moment.

Following Heidegger's removal of the modernist oppositions between subject and object, language and outside world, Merleau-Ponty's account states that a gesture is not a representation, or is not only such: it is a form of being. Gestures (and words, etc., as well) are themselves thinking in one of its many forms – not only expressions of thought, *but thought, i.e., cognitive being, itself*. To the speaker, gesture and speech are not only 'messages' or communications, but are a way of cognitively existing, of cognitively being, at the moment of speaking.

The speaker who creates a gesture of Sylvester rising up fused with the pipe's hollowness is, according to this interpretation, embodying thought in gesture, and this action – thought in action – was part of the person's being cognitively at that moment. Likewise the woman who gestured a sudden transformation with "gefunkt" and IW in his rotating metaphor of the 'getting into' process that he was undergoing. To make a gesture, from this perspective, is to bring thought into existence on a concrete plane, just as writing out a word can have a similar effect. The greater the felt departure of the thought from the immediate context, the more likely is its materialization in a gesture, because of this contribution to being. Thus, gestures are more or less elaborated depending on the importance of material realization to the existence of the thought. We observe the same elaboration of gesture in proportion to the importance of materialization in IW as well, and this is the final step of demonstrating the utter normality of his gestures of the 'throw-away' type.

Our second phenomenological excursion into the nature of speech and gesture concerns the notion of 'co-expressive non-redundancy', which was used to signify the convergence of two different modes of semiosis, the analytic/combinatoric verbal mode and the global/synthetic gestural mode, to represent one event (Sylvester climbing up inside the pipe) at the same time. An investigation of the concept of 'co-expressiveness' will shed light on how to interpret its non-redundancy.

---

[16]Merleau-Ponty's quotation is from Gelb and Goldstein (1925, p. 158).

[17]We are indebted to Jan Arnold for this quotation.

How is speech-gesture synchrony attained? Because the use of language and gesture *is* the speaker's taking up of a position in the world, *is* the speaker's way of cognitively being, the perfect synchrony of the different aspects of the speaker's expressive bodily behavior becomes self-evident.[18] As scientists we notice how well speech and gesture are attuned and how they break down together, but this is because we, as heirs to the modernist ideal of universal doubt, implicitly *first* take both linguistic modes as belonging to a different system, as having a life of their own, and *then* wonder how synchrony might be attained. In a framework, however, which takes cognitive being and the bodily expression of linguistic meaning to be one and the same, co-expressiveness becomes equal to bodily expressiveness in general. As we said, embodiment in the phenomenological sense is an a priori fact, and from this naturally follows that co-expressiveness of speech and gesture is a necessary given. Linguistic multimodality is the origin of meaning itself, and therefore the different modes are co-expressive.

What does this tell us about the 'non-redundancy' of the co-expressiveness? The very appearance of the concept of 'redundancy' in a discussion of linguistic multimodality belongs to a minimalistic framework in which the verbal is seen as the fundamental carrier of linguistic meaning and gesture as an additional mode (an "external accompaniment" of speech). When we ask the question "Why do we gesture?" we picture a still body which in the first place is capable of verbally expressing itself and which, in a linguistic event, may opt for the adding of gesture. Instead, if we take our active embodied existence as a given, we could also ask the question "Why wouldn't we gesture?" and picture a body engaged in the world, for which it is only natural to use its full capacities of expression. In this sense, what was 'redundant' not only becomes 'non-redundant', but even 'obligatory': using *all of your body* to convey linguistic meaning is standard practice.[19] Susan Goldin-Meadow (1999) has found that use of gesture reduces cognitive burden on the part of the hearer as well as on the part of the speaker, and as such a combination of speech and gesture makes the intended meaning more easily understandable (instead of soliciting the heightened cognitive activity which we would expect from

---

[18]This is Gallagher's point with respect to IW when he states that the timing of his gestures vis-à-vis his speech acts remains intact because "[t]he co-expressiveness of the two modes (gesture and speech) contribute to their synchronization." (2005:113)

[19]As this is a chapter on IW and gesture, we have focused on the manual modality. However, Merleau-Ponty's use of the term '*la geste*' cannot be unequivocally translated into 'gesture'. *La geste* refers to any aspect of the body deployed to convey meaning. But of course, because Merleau-Ponty's phenomenology of language is one about bodily expression in general, anything said there holds for the manual modality too. Historically, manual gestures have been the principal focus of observation (there may be evolutionary reasons to expect the hands to be primary) but studies, especially recent ones, have included the head (McClave et al. 2008), gaze (McNeill et al. in press) and vocal gestures (Okrent 2002) within a single framework of semiosis. These can be powerfully unified with the conception that linguistic meaning is obligatorily conveyed with all the body in unison (and that it is the suppression of elements that is exceptional).

an increase of contextual information). Our phenomenological framework can easily accommodate these findings because if we describe linguistic action in terms of a speaking subject making a contribution to being, using *more* co-expressive modalities will bring *more* of the same meaning about and for a listener it will be harder not to get what is expressed, as all bodily signs point into the same semiotic direction.

To end this section we will apply phenomenological philosophy to understand better the distinction between IW's 'throw-aways' and his 'constructeds'. Recall that his constructeds were fewer in number, were isolated, performed one by one and in a self-conscious manner. On the other hand, he produces his throw-aways with ease, though with some topokinetic problems. In a sense, by making this distinction, IW summarizes the whole point about the impossibility for third-person empirical-scientific approaches to fully capture the nature of gesture and for first-person phenomenological approaches to say anything objectively valid about gesture. When IW is unaware of his perfectly synchronized gesturing (when he is producing what he calls 'throw-aways'), he is immersed into the first-person point of view and he uses his whole body to convey his intentions. He bodily expresses his cognitive being at that time. However, when he is constructing his constructeds, he takes the third-person stance of the scientist who knows gesture from experience, but is unable to fully reconstruct it from a neutral point of view. He consciously divides his utterances and hand movements up by objectifying their features and then tries to attain synchrony. His cognitive being at that time (trying to control his hand moves) clashes with what he is trying to express with his hands (whatever the conversation is about). Co-expressiveness breaks down – and so does synchrony.

To Sum Up

To sum up this chapter we can ask: does IW show growth points; do his gestures act as material carriers; and do his meanings, in Merleau-Pontian fashion, inhabit them? IW's own distinction between 'constructed' and 'throw-away' gestures is critical at this point. His 'throw-aways' are indistinguishable from the gestures of unaffected speakers. That is, they comprise growth points with simultaneously encoded co-expressive linguistic content, to jointly differentiate what is newsworthy in context; offer the benefits of material carrierhood; and are inhabited by positions in his world of meanings. IW's very lack of awareness of them suggests this status. Unawareness is to be expected of positions in the world of meanings, and in this respect gestures are no different from most spoken words, of which, qua words, we are also usually unaware as we use them. The occurrence of this complex of processes in IW, despite deafferentation and his reworking of motion and control, suggests the existence of a thought-language-hand link in the human brain, the inheritance of us all, that survived his neuronpathy.

References

Bayne T (2004) Closing the gap? Some questions for neurophenomenology, Phenomenology and the Cognitive Sciences, 3, 4, 349–364

Braddock G (2001) Beyond reflection in naturalized phenomenology. Journal of Consciousness Studies 8(11): 3–16

Cole J (1995) Pride and a daily marathon. MIT Press, Cambridge, MA

Duncan S (2006) Co-expressivity of speech and gesture: manner of motion in Spanish, English, and Chinese. In: Proceedings of the 27th Berkeley Linguistic Society Annual Meeting, [Meeting in 2001.] Berkeley Linguistics Society, University of California, Berkeley, Department of Linguistics, 1203 Dwinelle Hall, Berkeley, CA, pp 353–370

Firbas J (1971) On the concept of communicative dynamism in the theory of functional sentence perspective. Philologica Pragensia 8:135–144

Gallagher S (2005) How the body shapes the mind. Oxford University Press (Clarendon), Oxford

Gelb A, Goldstein K (1925) Über Farbennamenamnesie. Psychologische Forschung 6:127–186

Goldin-Meadow S, Nusbaum H, Garber P, Church RB (2001) Transitions in learning: evidence for simultaneously activated hypotheses. J Exp Psychol Hum Percept Perform 19:1–16

Goldin-Meadow S (2003) Hearing gesture: how our hands help us think. Harvard University Press, Cambridge, MA

Goldin-Meadow S (1999) The role of gesture in communication and thinking. Trends Cogn Sci 3(11):419–429

Iverson JM, Goldin-Meadow S (1997) What's communication got to do with it? Gesture in congenitally blind children. Dev Psychol 33:453–467

Kendon A (1988) How gestures can become like words. In: Poyatos F (ed) Cross-cultural perspectives in nonverbal communication. Hogrefe, Toronto, pp 131–141

Mayberry R, Jaques J (2000) Gesture production during stuttered speech: insights into the nature of gesture-speech integration. In: McNeill D (ed) Language and gesture. Cambridge University Press, Cambridge, pp 199–214

McClave E, Kim H, Tamer R, Mileff M (2008) Linguistic movements of the head in Arabic, Bulgarian, Korean, and African American Vernacular English. Gesture

McNeill D, Duncan SD (2000) Growth points in thinking for speaking. In: McNeill D (ed) Language and gesture. Cambridge University Press, Cambridge, pp 141–161

McNeill D, Duncan S, Cole J, Gallagher S, Bertenthal B (2008) Growth points from the very beginning. Int Stud 9:117–132

McNeill D, Duncan S, Franklin A, Kimbara I, Parrill F, Welji H. In press. Mind-merging. In: Ezequiel M (ed) Expressing oneself/expressing one's self: communication, language, cognition, and identity: a festschrift in honor of Robert M. Krauss. Taylor & Francis

McNeill D (1992) Hand and mind: what gestures reveal about thought. University of Chicago Press, Chicago, IL

McNeill D (2005) Gesture and thought. University of Chicago Press, Chicago, IL

Merleau-Ponty M (1962) Phenomenology of perception (trans: Smith C). Routledge, London

Müller C (forthcoming) Metaphors – dead and alive, sleeping and waking. A dynamic view. University of Chicago Press, Chicago, IL

Nobe S (2000) Where do *most* spontaneous representational gestures actually occur with respect to speech? In: McNeill D (ed) Language and gesture. Cambridge University Press, Cambridge, pp 186–198

Okrent A (2002) A modality-free notion of gesture and how it can help us with the morpheme vs. gesture question in sign language linguistics. In: Meier RP, Cormier K, Quinto-Pozos D (eds)

Modality and structure in signed and spoken language. Cambridge University Press, Cambridge, pp 175–198

Overgaard M (2004) On the naturalising of phenomenology, Phenomenology and the Cognitive Sciences, 3, 4, 365–379

Overgaard M, Fehl K, Mouridsen K, Cleeremans A (2008) Seeing without seeing? Degraded conscious vision in a blindsight patient, PLoS ONE, 3,8, 1–4

Parrill F (2007) Metagesture: an analysis of theoretical discourse about multimodal language. In: Duncan SD, Cassell J, Levy E (eds) Gesture and the dynamic dimension of language. John Benjamins, Amsterdam

Ramachandran VS, Blakeslee S (1998) Phantoms in the brain: probing the mysteries of the human mind. William Morrow, New York

Rieber RW, Carton AS (eds) (1987) The collected works of LS Vygotsky, vol 1: Problems of general psychology. Including the volume THINKING AND SPEECH (intro. and trans: Norris M). Plenum, New York

Roy JM, Petitot J, Pachoud B, Varela FJ (1999) Beyond the gap: an introduction to naturalising phenomenology, in Petitot et al. (eds.): Naturalising Phenomenology, Princeton University Press

Severance E, Washburn MF (1907) The loss of associative power in words after long fixation. Am J Psychol 18:182–186

Slobin DI (1987) Thinking for Speaking. In: Aske J, Beery N, Michaelis L, Filip H (eds) Proceedings of the thirteenth annual meeting of the berkeley linguistic society. Berkeley Linguistic Society, Berkeley, pp 435–445

Varela FJ (1996) Neurophenomenology: Amethodological remedy to the hard problem. Journal of Consciousness Studies 3(4): 330–350

Part VIII
Pathologies

Phenomenology and Psychopathology

Thomas Fuchs

Introduction

As the systematic project of investigating the structures of subjective experience, phenomenology may also be considered the foundational science for psychopathology. Though it methodically suspends any assumptions about causal explanation, it provides a rich framework for the analysis of subjectivity and its disturbances in mental disorders, thus also leading to testable hypotheses about the underlying neurobiological mechanisms. Whereas the first movement of phenomenological and existential psychiatry – mainly derived from European, particularly German and French sources – came to a certain conclusion in the 1970s (marked by Spiegelberg's synopsis in 1972), the last two decades have seen an international revival of phenomenological psychopathology which also entered into a constructive dialogue with cognitive neuroscience (Parnas and Bovet 1995; Mishara et al. 1998; Fuchs 2002a).

Present phenomenological psychopathology has gained new ground by emphasizing the roots of mental illness in the patients' prereflective experience. Drawing on the advances of phenomenological research in general, it relates psychopathology to the basic structures of consciousness such as self-awareness, embodiment, spatiality, temporality, intentionality, and intersubjectivity. In order to investigate these dimensions, the phenomenologist will start with questions such as:

– What is it like for the patient to be in a certain mental state (e.g. to feel depressed or to hear voices)? What is the personal meaning of that state?[1]

T. Fuchs
Department of Psychiatry, Center for Psychosocial Medicine, University of Heidelberg

[1]To take an example mentioned by Stanghellini (2007): What exactly does a patient mean e.g. when he says 'I feel depressed'? – Some patients may use the word 'depressed' to describe themselves as feeling sad and downhearted, discouraged by a setback or another adversity, corresponding to a reactive depression. Others may use it to mean that they feel dull, empty, dysphoric and bored, as is often the case in Borderline patients. Others may denote that they are unable to feel, that they have lost the affective resonance with others, like being petrified – corresponding to the 'feeling of loss of feelings' in endogenous depression. Finally, some patients may try to convey their sense of inner void, lack of inner nucleus or of identity, feelings of being anonymous or non-existent, as occurring in the prodromal phases of schizophrenia..

– How does the patient experience his or her world? How does he or she express, move, and define space as an embodied subject?
– What is the subject's experience of existential time? Is there a sense of continuity over time, or are there breaks or fadings of self-awareness?
– Does the patient feel effective as an agent in the world, or rather as only being exposed to the world?
– Is there a tendency to take an external perspective to one's body, actions, and self? Do the knowing and the feeling subject coincide or diverge?
– How is the patient's ability to empathize with others, to take their perspective? How does he/she experience his or her relationships?

In this way starting with first-person accounts, phenomenology proceeds to the constitutive processes that build up subjective experience, such as the formation of perceptual meaning, temporal continuity or implicit bodily action. This also allows for the detection of the critical points where the constitution of self and world is vulnerable and open to deviations or derailments. Particularly in psychotic disorders, the subject may lose its ground in the lived body as the ensemble of dispositions and habitualities, its anchoring in temporal continuity and in intersubjective common sense (Stanghellini 2004; Fuchs 2005a). On the other hand, despite the erosion of the constitutional processes, the patients still strive for a coherent world view, though this may only be possible in the form of delusion or autistic withdrawal. Accordingly, the phenomenological psychopathologist also explores the modes by which the patients try to make sense of the basic disturbances and to re-establish some form of coherence.

Phenomenology does not consider subjectivity as an object to be described but as a medium allowing the world to manifest itself. Therefore it aims at grasping not the content or object, but rather the *form and structure* of conscious experience. It is likely that the altered form is, pathogenetically speaking, closer to the biological substrate, whereas content is more contingent or idiosyncratic. However, phenomenology does not consider symptoms of mental illness in isolation, i.e., as disconnected manifestations of localized brain dysfunctions, but in relation to the subject and the whole of consciousness in which these symptoms emerge. Thus, the phenomenological approach creates an *intermediate level* that relates the level of molecular dysfunctions as studied by experimental neuropsychology (e.g. deteriorated working memory, executive control functions, attentional disturbances, etc.) to the molar level of descriptive psychopathology and its nosological syndromes. The micro-dysfunctions may be integrated into a comprehensive account of altered self-experience.

This also applies to the level of diagnosis: Diagnostic entities are seen by phenomenology not as statistically relevant clusterings of symptoms, but rather as certain *typical modes of human experience and existence*, reflected in their invariant phenomenological structures – independent from nosological classification or epidemiological data on comorbidity. What phenomenology is looking for instead are the "psychopathological organizers" that connect the single features, for example, affective depersonalisation in melancholic depression or autism in schizophrenia.

Thus it helps to define mental disorders on the basis of their structural experiential features, linking apparently disconnected phenomena together.

The basic assumption guiding the phenomenological approach is that human subjectivity and intersubjectivity are characterized by an inherently vulnerable structure which may lead to the derailments, alienation and disintegration found in psychopathology. The multiple synthetic processes necessary for constituting the sensory-motor and temporal unity of conscious experience are amenable to a variety of possible disturbances. Moreover, self-alienation as the hallmark of mental illness is prefigured in the dialectical, precarious and unstable condition of the human subject which by being self-conscious relates to itself, but always escapes itself at the same time. This creates a tension or fracture which bears the germ of alienation, and which has to be constantly bridged by the subject's engagement in the world. Existence as self-transcendence is the continuous movement driven by this basic instability, by a self-alienation *in statu nascendi,* as it were, that we have to overcome continuously. As Karl Jaspers put it in his *General Psychopathology*: "For man, his uncompleteness, his openness, his freedom and illimitable possibility itself becomes the cause of illness."[2]

In this chapter, I will consider three aspects of human subjectivity that are particularly vulnerable to disintegration and alienation: (1) embodiment with its basic antagonism of subject-body and object-body, (2) temporality with its antagonism of past- and future-orientation, and (3) intersubjectivity with the complex dialectics of perspective-taking and self-other-distinction. As a framework for these analyses, I will first give a short phenomenological account of self-experience on the basic, pre-reflective and on the extended, reflective or narrative level.

Dimensions of Self-experience

Since all major psychiatric disorders involve a more or less pronounced disturbance of the self in its relation to the world, the phenomenology of self and self-awareness is of essential importance for psychopathology as well. Referring to the concepts currently discussed in phenomenology, developmental psychology and cognitive neuroscience, we may distinguish the *minimal or core self* from the *extended or narrative self* (Damasio 1999; Gallagher 2000a; Rochat 2004; Zahavi 2005).

(1) The *minimal self* is characterized by an implicit, prereflective self-awareness that is present in every experience without requiring introspection. Thus, any sensation, any perception or action directed towards an object implies a tacit self-awareness; it is given immediately, non-inferentially as mine. This first-personal givenness of all experience may be regarded as a general medium in which specific modes of experience are articulated. As the most basic form of selfhood, it may also be called 'mineness' or *ipseity* (from the Latin *ipse* = 'self' or 'himself'; Henry 2000; Zahavi 1999). Ipseity is preserved even

_____

[2] "Dem Menschsein ist seine Unfertigkeit, seine Offenheit, seine Freiheit und seine unabschließbare Möglichkeit selber Grund eines Krankseins" (Karl Jaspers, *Allgemeine Psychopathologie*, p. 8).

when autobiographical memory is lost, as in amnesia or dementia, or when a long-term sense for the future is missing, as in certain frontal brain damages.

On the other hand, the basic self should not be conceived as an abstract, disengaged ego, but as involving the dimensions of *self-affection, embodiment* and *temporality*. Ipseity or 'mineness' is bound to the background feeling of the body, mediated by proprioceptive and kinaesthetic awareness, and implies a basic self-affectability or auto-affection (Henry 2000). Moreover, it involves the sensory-motor relation to the world mediated through the body with its particular constitution and its habitual background capacities. By being embodied and thus structurally coupled with a complementary environment, the basic self becomes an "ecological self" (Neisser 1988). It is embedded into its *lived space* and *lived world* which presents itself as a field of possibilities, affordances, barriers or obstacles (Fuchs 2007a). Finally, prereflective self-awareness also implies a basic *temporal continuity* as analyzed by Husserl in his phenomenology of inner time consciousness: The continuous intertwining of succeeding moments by 'retentions' and 'protentions' includes an intrinsic awareness of my ongoing experience *as* mine (see Section on Disturbances of Temporality below). Thus, the phenomenological analysis of the temporal structure of consciousness is capable of accounting for "… self-identity through time, without actually having to posit the self as a separate entity over and above the stream of consciousness" (Gallagher and Zahavi 2005).

(2) The *extended or narrative self* begins to shape in the second year of life. It is based on a number of emerging capacities that are closely interrelated:

   – The capacity for a higher-order awareness of one's conscious states, i.e. introspective or *reflective self-consciousness*
   – The capacity to understand others as intentional agents and to take their perspective, i.e. *self-transcendence*
   – The capacity to understand and issue verbal reports about one's own or others' feelings, thoughts and intentions, i.e. *narrativity*
   – The capacity to form a conceptual and autobiographical knowledge of oneself, i.e. a *self-concept*

The extended self emerges in the course of early socialisation, depending on the acquisition of autobiographical memory, concepts and language. Its fundamental structure is intersubjective and reciprocal: It is constituted through the ongoing relation to others, as the 'social self' or *'me'* conceived by G. H. Mead (1924), which includes seeing oneself 'in others' eyes', internalizing their attitudes toward oneself and gradually adopting the roles offered by the community. Taking the perspective of others implies a shift from ego-centric to allo-centric space and a concept of oneself and others as intentional agents who are responsible for their actions. This is not only a cognitive achievement, but gives rise to a number of 'self-reflective emotions' such as shame, embarrassment, feelings of guilt or pride which all depend on the internalized, evaluating 'gaze of the other' (Seidler 2000; Fuchs 2002b). Any narrative only makes sense for a real or an implicit other as well. There is an inner witness in most of our actions and intentions, to whom we

could give an account of what we did, and justify what we are doing. In relating to herself, the person lives immersed in narratives, at the intersection of different stories already begun but not yet completed.

Despite this complex and dialectical structure, the extended self always remains based on prereflective self-awareness: Only a being with the constant sense of mineness is able "to form concepts about herself, consider her own aims, ideals, and aspirations as her own, construct stories about herself, and plan and execute actions for which she will take responsibility" (Gallagher and Zahavi 2005). Disturbances of basic self-awareness are therefore bound to affect the extended self as well. Both dimensions of self-experience, however, are not present a priori, but have to be constituted by a complex interaction of biological, psychological and social processes. The coherence of the self is a constant achievement of self-constitution and self-affirmation – an achievement which may be disturbed in manifold ways.

Disturbances of Embodiment

Corresponding to the levels of basic and extended self-experience, phenomenologists usually distinguish between (a) the body that I prereflectively live as an incarnated subject, i.e. the *subject-body (Leib),* and (b) the physical body that I can perceive or that is perceived by others, the *object-body (Körper).*[3]

(a) Originally, the body functions in a tacit or implicit way, as the very center and medium of subjective experience. It constitutes the zero-point that permits my perceptual view on the world, while it is itself not perceived; it operates in every action and interaction with others, without requiring explicit attention. This operative intentionality of the body (Merleau-Ponty 1945) is based on the "passive syntheses"[4] of sensory-motor functioning that link elements of perceptions and movements into higher-order schemas or *Gestalten*, thus forming the background texture of the field of experience. In this way, my body represents my situatedness, my perspective as well as my potentiality within the world.[5]

(b) Normally, the body tends to efface itself in our world-directed activity (Leder 1992). However, it appears as an object of conscious attention particularly when it is inadequate for a task to be performed, be it by a lack of capacity, fatigue or illness; and whenever it becomes an object for others to whom I feel exposed. In these cases, the body's performance is made explicit and may often be disturbed. Thus the body has a double or ambiguous experiential status: both as a 'lived body', implicit in one's ongoing experience, and as an explicit, physical or objective body. The subject-body means my openness to a future, my general capacity or potentiality, and constantly surpasses the object-body

---

[3] See, for example, Husserl (1973, p. 57), Merleau-Ponty (1945), Plessner (1981).

[4] 'Passive synthesis' is the term Husserl used to denote a synthesis that is not brought about by a conscious act but is intrinsic to consciousness itself (Husserl 1966).

[5] The German term '*Befinden*' ('situatedness', 'condition', 'well-' or 'ill-being') aptly denotes the background feeling of the body as well as 'finding oneself in a certain situation'.

which is only present in retrospection, as my body-past. An ongoing oscillation between these two bodily modes constitutes a fluid and hardly noticed foundation of all experiencing. The philosopher Helmut Plessner (1981) coined the term *'excentric position'* to characterise the ambiguous status of the embodied human person between being inside of her body, in the center of her world, and being outside of it, in reflective distance from pure centrality.

A closely related distinction is the one between *body schema* and *body image*. The body schema means a complex interplay of sensory-motor systems (e.g. the visual, vestibular, proprioceptive, kinesthetic system) regulating bodily posture and movement in relation to the environment – in short, the non-conscious performance of the body. The body image, on the other hand, signifies a "system of perceptions, emotional attitudes, and conceptual beliefs that pertain to one's own body" (Gallagher and Vaever 2004, p. 119 2004a, p. 119), or the conscious awareness of one's body. The one corresponds to an *implicit and egocentric*, the other to an *explicit and allocentric* perspective.

On this basis, disturbances of embodiment may be classified

(a) As primarily affecting the *subject-body* or prereflective embodied sense of self
(b) As being related to the *body image or explicit body awareness*

In what follows, I will give an overview on different types of disturbances of embodiment.

Disturbances of the Subject-Body

Schizophrenia as a Disembodiment

Current neuropsychological theories attribute the core disturbances in schizophrenia to higher order cognitive processes such as "theory of mind" and self-monitoring or "meta-representation" (Frith 1992, 2004). In contrast, phenomenological approaches, in accordance with research on basic symptoms by Huber and Klosterkoetter (Gross et al. 1987), locate the essence of the schizophrenic syndrome in disturbances of prereflective self-awareness and embodiment. This includes: (1) a weakening of the basic sense of self or ipseity (Sass and Parnas 2003) and (2) a disruption of implicit bodily functioning in the dimensions of both perception and action. As a result, the prereflective, embodied and practical immersion of the self in the world is fundamentally disturbed.

1. According to the phenomenological theory put forward by Parnas and Sass, schizophrenia involves a diminishment of basic self-awareness, a feeling of a pervasive inner void or lack, and an increasing anonymity and depersonalization of the field of awareness (Parnas and Sass 2001, Sass and Parnas 2003). The loss of vital contact with reality may be expressed in complaints about a certain unclarity or opacity of consciousness ("living like in a fog", "feeling surrounded

by invisible walls"), but also in a general feeling of being alien to the world. Disorders of basic self-awareness have recently come to be explored in detail by means of an extensive, phenomenologically based interview (EASE, Examination of Anomalous Self-Experience, Parnas et al. 2005).

2. The disturbance of ipseity is accompanied by a loss of automatic processing on the level of 'passive syntheses', leading to an increasing fragmentation of perceptual and motor Gestalt schemas, and to a 'pathological explication' of the implicit functions of the body (Sass 2000; Fuchs 2005a). Thus, otherwise tacit sensory-motor processes become available for introspection. I will describe this in more detail.

In *perception*, the dissolution of passive syntheses normally organising the perceptual field manifests in an impaired capacity to recognize familiar patterns or *Gestalten* (Wiggins et al. 1990; Wiggins and Schwartz 2007). Patients often experience an overload of details separated from their situational context, without grasping the scene's overall meaning. With growing alienation, even *the act of perceiving itself* may come to awareness; then the patients are like the spectators of their own perceptive processes.[6] This disembodiment and alienation of perception turns the objects into mere appearances or phantoms; hence the artificial, enigmatic alteration of the environment experienced especially in the early stages of psychosis (Fuchs 2005a). At the same time, new saliencies may emerge, i.e. expressive qualities, strange features of persons and faces, or hypersignificant objects standing out from the incoherent background. Over the course of time, these noncontextualized fragments are reorganized by emerging delusions that provide a new but rigid coherence of the perceptual field by sacrificing some features while preserving others. In experimental studies, schizophrenic patients show an impairment of sensory processing, in particular deficits in the grouping of stimulus elements into coherent object representations; this is reflected in reduced phase synchrony of neural responses (EEG gamma-waves; Uhlhaas and Mishara 2007). Thus, we find a convergence of evidence from phenomenology and experimental research.

A similar alienation concerns bodily functioning in *movement and action*. Schizophrenic patients often speak of a split between their mind and their body, of feeling detached from their lived performance like a machine or a robot. In particular, they may experience a disintegration of habits or automatic practices, a "disautomation". Instead of simply dressing, driving, walking, etc., they have to prepare and produce each single action deliberately, in a way that could be called a "Cartesian" action of the mind on the body. Thus, the units of meaningful actions are fragmented, resulting in a pathological explication and hyperreflexive awareness of normally tacit aspects of everyday behaviour (Sass and Parnas 2003). In advanced stages, the *sense of agency* for one's actions (i.e. the sense that I am the one initiating the movement) may be disrupted, finally leading to delusions of alien control. On the neurological

---

[6] Thus, a patient reports "I become aware of my eye watching an object" (Stanghellini 2004, p. 113), or "I saw everything I did like a film-camera" (Sass 1992, p. 132).

level, this may be caused by a dysfunction of forward comparator processes in premotor and prefrontal cortices as well as in the supplementary motor area which normally serve as a neuronal 'who' system for distinguishing self-initiated actions from actions observed in others (Georgieff and Jeannerod 1998). However, the *sense of mineness or ownership* for these actions is preserved, since they are experienced as still belonging to one's own body. The sense of ownership probably depends on sensory feedback mechanisms that are intact in schizophrenia (Gallagher 2000b, 2004a).

In sum, schizophrenia implies a disembodiment of the self in the sense of losing one's habitual bodily performance, and with it the prereflective, questionless being-in-the-world that is mediated by the body. The basic disturbance may also be regarded as a loss of "common sense", i.e. the tacit knowledge and familiarity with the world and with others (Polanyi 1967; Blankenburg 2001; Stanghellini 2007). Common sense is the way that past experience implicitly informs our current perceptions and actions. It provides a fluid, automatic and context-sensitive pre-understanding of everyday situations, thus connecting self and world through a basic habituality and familiarity. In schizophrenia, patients experience a "loss of natural self-evidence" (Blankenburg 1971), a lack of tacit attunement to other people and situations. They report feeling isolated and detached, unable to grasp the "natural", everyday meanings of the common world. Thus, the relationship of self and world is in constant need of being reconstructed by deliberate efforts, leading to the growing perplexity and hyperreflexive ruminations that are found in schizophrenic patients (Sass 2000).[7]

Melancholic Depression as "Hyperembodiment"

A different disturbance of embodiment is found in melancholic depression. Here, the body loses the fluidity and mobility of a medium and turns into a heavy, solid body which puts up resistance to the subject's intentions and impulses. Its materiality and weight, otherwise suspended in everyday performance, comes to the fore and is experienced as a leaden heaviness, oppression and ridigity (e.g. a feeling of a tyre around the chest, pressure in the head, or as general tightness and anxiety). Thus, instead of giving access to the world, the body stands in the way as an obstacle, separated from its surroundings: The phenomenal space is not embodied any more. However, this is not only due to psychomotor inhibition (as, for example, in Parkinson's disease). Rather, the *conative dimension* of the body, i.e. its seeking for satisfaction, is missing. Normally, it is this dimension which opens up the peripersonal space as a realm of possibilities, 'affordances' and goals for action. In depressive patients, however, drive

---

[7] On the neurological level, these analyses may be correlated with concepts of schizophrenia as a neurodevelopmental disorder (Marenco and Weinberger 2000). A lack of intermodal integration due to faulty maturation of cortico-cortical connectivity could result in an impaired development of the "ecological self" and its perceptual, cognitive and emotional ties with the natural and social environment (Parnas et al. 1996, 2002).

and impulse, appetite and libido are reduced or lost, no more disclosing potential sources of pleasure and satisfaction. Confined to the present state of bodily restriction, the depressive person cannot *transcend* her body any more. With growing inhibition, her sensory-motor space is restricted to the nearest environment, culminating in depressive stupor. Thus, melancholia may be described as a reification or 'corporealization' of the lived body, a 'hyperembodiment' (Fuchs 2002a).

At the same time, there is a more subtle loss of the bodily resonance or affectability that mediates emotional experience and the affective attunement with others. Since the corporealized body loses its capacity of emotional resonance, the patients feel inanimate, detached from their emotions, and complain of a "feeling of not feeling". They are no longer capable of being moved and affected; the attractive and sympathetic qualities of their surroundings have vanished. Since loss of feeling means a diminished prereflective sense of self, *affective depersonalisation* is the clinical core-feature of severe melancholic episodes (Kraus 2002; Stanghellini 2004; Fuchs 2005a). In some cases, the depersonalisation culminates in the so-called nihilistic delusions or Cotard's syndrome (Enoch and Trethowan 1991). Patients then claim that they have already died, and their body has turned into a corpse. They may even deny their own existence or the existence of the world. This can be understood as a separation of the "pure", unaffected consciousness from the corporealized body, whose heaviness now changes to the opposite, i.e. a feeling of lightness or even to a complete loss of bodily sensations: Proprioception, taste, smell, the sense of warmth or pain may be missing. Thus, the sense of bodily ownership or auto-affection is severely disturbed, while the sense of agency is still present.

To summarize, the person affected by melancholia collapses into the spatial boundaries of her own solid, material body. Instead of transcending the physical body, she becomes completely identified with it: Unable to detach herself from the experience of bodily failure, she feels worthless, guilty or decaying. While this may be termed a 'hyperembodiment", in the nihilistic culmination of melancholia the self disconnects from the corporealized body; by this, however, it loses the sense of being alive. Similar types of disembodiment may occur in other depersonalization syndromes, where the affective sense of self is disintegrated and the body is experienced as an object among others.

Disturbances of the Body-Image (Object-Body)

In contrast to disturbances of prereflective embodied self-awareness, other kinds of disorders originate from the explicit relation of the subject towards his or her body, i.e. from *body-image*. Thus, the situation of being shamefully exposed to others' gazes may give rise to disorders such as social phobia or body dysmorphic disorder, while a more complex combination of affective, conceptual and social aspects of the body image is involved in hypochondriasis, somatoform disorders and anorexia nervosa. As examples, we will have a closer look at *body dysmorphic disorder* and *anorexia*.

Body Dysmorphic Disorder

The potential self-alienation that arises from becoming aware of oneself in others' eyes has been famously analyzed by Sartre (1943). Exposed to their gazes, my prereflective embodied being is turned inside-out, as it were, and my world is decentralized. The lived-body becomes a body-for-others, i.e. an observed, unprotected or denuded object-body, exposed to potential evaluation or rejection. This is the origin of several self-reflective emotions, in particular, of embarrassment and shame. In shame, one is painfully affected by centripetal directions, that is, by piercing gazes or pointing fingers which one tries to escape in vain. The ashamed person becomes the focus of felt attention, paralyzed and reified through the others' gazes, and loses her natural centrality and self-confidence. Thus, shame is of particular importance for the pathologies of the body image (Fuchs 2002b). They typically manifest themselves for the first time in adolescence, when the body changes and gains a new external aspect through sexual maturation.

Body dysmorphic disorder is characterized by overvalued fears of an assumed ugliness or deformity of actually inconspicuous body parts. The patients complain of a huge nose, a misshapen form of the mouth or other parts, excessive hair in the face, swelling or reddening of the complexion, etc. Often the body part concerned is felt as prominent and as bigger than before. Thus, the reification of the body through the other's gaze focuses on one part as *'pars pro toto'*. Body dysmorphic shame is increased by the patient's egocentristic and hyperreflexive stance. Fear of visual exposure and feelings of being constantly observed, stared or laughed at by the others may culminate in paranoid ideas of reference (Pinto and Phillips 2005). Cosmetic surgery is often sought, but as a rule does not alter the severe lack of self-esteem the disorder is based upon. The 'body-for-others' now dominates the lived-body and leads to sociophobic avoidance.

Patients with body dysmorphic disorder are mostly characterized by sensitive, dependent, ambitious or narcissistic tendencies (Phillips 2000). They are especially threatened by set-backs, humiliations or failures in the interpersonal sphere. Thus, the pathology of the body image is caused by a disturbance in their social relationships which, however, remains hidden to themselves. The patient's bodily appearance stands only vicariously for an insufficiency of his or her basic self-esteem (Phillips et al. 2004; Buhlmann et al. 2007). Feeling his own self-devaluation in the other's gazes, the patient is overwhelmed by their perspective on himself, unable to gain an independent point of view. The lived body becomes conscious as corporeal body, and with it, the patient's thinking constantly revolves around the body part. Since self-awareness is thus fixed on the isolated body-object, the vicious circle of reification and shameful self-awareness can no more be interrupted.

Anorexia Nervosa

In anorexia nervosa, the dialectic between *being* a body-subject and *having* one's body as an object becomes the core of the disorder. The anorectic patient refuses

the dependence on the prereflective, natural body with its uncontrollable and obscure becoming, its threatening impulses and cravings – in particular hunger and sexual desire – which may only be quenched by external supply. This self-willed, appetent body now becomes an alienated, repulsive object that may even arouse disgust. Swallowing food is experienced as the incorporation of a foreign substance that stuffs the body and causes it to bulge. The anorectic patient rejects her hungry, dependent body, but also denies the maturation of her female, swelling and sexual body. Gaining independence from it, and turning it into an object of control and mastery, becomes a source of grandiose triumph (Walter et al. 2007). Thus, the implicit sense of bodily ownership is replaced by an explicit appropriation of the body aimed at perfect control and maximal suppression of need and desire.[8]

Though distortions of the body image, such as feelings of fatness and unattractiveness, overestimation of body size and weight[9], play an important role, these external aspects of the body image constitute only ostensible motives, not the primary source of the addictive starvation. Rather, seeking to compensate for a lacking sense of identity and autonomy, the patients gain a feeling of accomplishment by rigorously subjugating and modelling their body: "I do not feel hunger any more; I am self-sufficient and don't need anything from outside" (Dignon et al. 2006). Similarly, the cessation of menstruation and loss of sexual desire renders the anorectic independent both from her body and from others. Thus, the body expresses a particular disturbance of the patient's relationships: her appearance and comportment cannot fail to reject those around her. "To look at her is awkward; she looks like a dying person. To hug her is uncomfortable; she feels hard and skeletal" (Jacobson 2007). Moreover, to eat with an anorectic is alienating and tormenting. In withdrawing from eating, she is also expressing a rejection of sociability – literally, of sharing bread with others.

In sum, for the anorectic patient, being in control over her body becomes synonymous with being in control of her life. By turning the fluid lived body into a rigid object body she also blocks the everyday intercorporeal contact with others. Certainly, cultural influences on the body image, i.e. the promotion of thinness as the ideal female form in Western nations, play an important role. However, for the patients slenderness is not aimed at sexual attractiveness, but rather constitutes an esthetic ideal of a seraphic, asexual and self-contained body. Thus, in anorexia the self is alienated from its prereflective embodied being; the inevitable ambiguity of being and having a body is turned into a sharp dualism, reminiscent of Platonic and Christian traditions of the body as "the soul's dungeon".

---

[8] "Anorexia" is derived from the Greek *"órexis"* which means "desire", "striving".

[9] Recent reviews of research suggest that the distortion is not a perceptual problem, but one of how the perceptual information is evaluated by the affected person (Skrzypek et al. 2001, Benninghoven et al. 2007).

Disturbances of Temporality

Basic and extended self-awareness are also associated with two different levels of phenomenological time: (a) a *basic or implicit temporality* of conscious experience, and (b) *explicit or autobiographical temporality*.

(a) Basic, prereflectively lived temporality is not a psychological category, but an intrinsic feature of conscious experience. Lived time runs with the movement of life, unfolding through the processes of embodied activity. Even in its most basic forms, consciousness is constituted as the duration or extension of awareness that spans succeeding moments and thus establishes a fundamental continuity. According to Husserl's analyses of 'internal time consciousness' (Husserl 1991), there is a constant 'passive synthesis' of an impressional present with its predecessors being retained (retention), and with its successors being intended at the same time (protention). The now does not exist in isolation but as the "temporal fringe" or "temporal field" spanned by retentions and protentions. Thus, a melody is not the mere succession or sum of single tones, but their integration into a temporal *gestalt*. This constant integration is equivalent to the continuity of the basic self: At the same time that I am aware of a melody, I am co-aware of my ongoing experience of the melody; self-awareness is implicit in my experience of intentional content (Gallagher and Zahavi 2005).

According to Merleau-Ponty (1960), this basic temporal structure constitutes the 'intentional arc' or the operative intentionality of embodied subjectivity. It is operant in every perception (as being based on comparison and detection of changes) as well as in every action (as being based on motivational and sensory-motor anticipation of the goal). Fuster has related this temporal integration to the functions of the prefrontal cortex comprising the tripartite functions of *working memory* (related to the past), *interference control* (related to the present) and *preparatory set* (related to the future). "Integration across time is a basic function of the prefrontal cortex and the basis of its cardinal role in the temporal organization of behaviour" (Fuster 2003). Thus, the readiness potential in the premotor cortex may be seen as a neural indicator of intended actions. The inhibitory interference control protects the *gestalt* of actions or behaviour from interfering influences or impulses. Obviously this is also necessary for any selective attention. Moreover, time estimation studies have pointed out that the distributed interaction of the prefrontal cortex with the cerebellum, the basal ganglia and the inferior temporal lobe is essential for temporal information processing in the brain (Vogeley and Kupke 2007).

(b) Whereas in prereflective (implicit) temporality future and past do not stand out against the pure presence of 'becoming', *explicit or autobiographical temporality* arises when the individual becomes aware of her own past and future, her origins and her finitude, and realizes that life is not merely given, but a life to be lead, or even a task to be performed. Explicit temporality is based on a sense of personal continuity over time, of being a person with certain abilities, character traits, interests, goals and convictions. It manifests

itself particularly in the ability to make plans and promises, to pursue one's goals even against resistance, and thus to project oneself into the future. It also implies the constant attempt at a meaningful integration of one's life history, using the narrative models and accounts prevalent in one's cultural environment (Taylor 1989; Philipps 2003). Thus, temporality, narrativity and coherence of identity are closely intertwined.

Disturbances of Basic Temporality

Disturbances of basic temporality may affect, on the one hand, the motivational and energetic dynamics of mental life. Affective disorders typically present alterations of the experienced velocity of time flow, with acceleration in mania and retardation in depression. Time estimation studies have mostly found a distension of subjective time experience in depression, and an abbreviation in hypomanic or manic episodes (Mezey and Knight 1965; Bech 1975; Kitamura and Kumar 1982; Münzel et al. 1988; Mundt et al. 1998). This corresponds to phenomenological conceptions of affective disorders as disturbances of self-temporalization (Tellenbach 1980; Fuchs 2001a).

Whereas the basic temporal structure is preserved in these disorders, schizophrenic experiences such as thought disorder, loss of agency or delusions of control may at least in part be explained by a disturbance and fragmentation of internal time consciousness (Fuchs 2000, 2007; Gallagher 2005; Vogeley and Kupke 2007). As is well known, schizophrenic patients show impaired attentional spans, difficulties in planning, initiating and coherently performing action or speech, as well as disruptions of experiential continuity. Moreover, disturbances of the temporal sequencing and synchronisation of cognitive, perceptive and motor functions have been described in schizophrenia, leading to the concept of "cognitive dysmetria" as a basic disturbance (Andreason et al. 1998). A hypothetical disconnection syndrome in cortico-cerebellar-thalamic circuits could result in a failing coordination of mental and motor activity. In phenomenological terms, these disturbances correspond to a fragmentation of the intentional arcs of thought, action and perception.

A closer analysis, based on Husserl's concepts, points to an impairment and intermittent failure of the *protentional* function of consciousness as the cause of this fragmentation (Gallagher 2000b; Fuchs 2000, pp. 144ff., 2007; Mishara 2007). Because schizophrenic patients still experience disrupted thoughts or actions as belonging to themselves, the retentional function which may be related to the sense of ownership seems to be effective. However, a weakening of protentions would not only lead to a reduced attentional span or thought disorder, but also to a disturbance of *agency*. Gallagher has proposed that agency is generated precisely in the processes that anticipate the thought or action. If these anticipatory, preparatory or protentional processes are disturbed, unfitting associations or movements may intrude, but can only be experienced in the retentional mode. The subject is then no more actively directed towards the future, but is left with focusing on what just turned up in his consciousness, or on the sensory feedback of his just-past

movement. This *delayed or retroactive consciousness* is captured in the following description of a schizophrenic patient:

> My feeling of experience *as my own experience* only appears a split second delayed (Parnas 2005, p. 245).

This kind of delay may be regarded as a decisive basis of psychotic self-disorders: Since the retentional-protentional structure implies a tacit self-awareness as well (see above), a fragmentation of this structure causes a disruption of self-continuity. Consequently, the unintended and unforeseen events of thinking or acting are experienced as alien to the self, i.e. as thought insertions, auditory hallucinations or alien control of action.

Neuropsychological evidence links these analyses to disturbances of working memory and of executive control functions mainly located in the prefrontal cortex (Fuster 1997; Harrington et al. 1998; Manoach 2003), in particular, with the timing or sequencing component of mental activity. Patients with schizophrenia show marked deficits of working memory which sometimes manifest themselves as formal thought disorders (Vogeley et al. 1999). Further, they may exhibit a retardation and disturbance of sequential finger movements (Jirsa et al. 1996), a reduced ability to discriminate stimuli in close temporal vicinity (Braus et al. 2001), and abnormally long latencies in estimating time intervals (Mishara 2007). This is in accordance with the phenomenological account: If events occur faster than anticipated, the protentional function would be 'overwhelmed', and perplexity would result when the patients try to interpret the meaning of what intrudes on them. Finally, the emergence of delusions may be regarded as the formation of a fixed framework that integrates the nontemporalized fragments in a preconceived, rigid schema of meaning, at the price of a circumscribed closure of the future as a dimension of open possibilities.

In sum, there is increasing evidence for a structural homology between phenomenology and cognitive neuroscience in time consciousness. Empirical data and theoretical models relate the postulated disturbance of time consciousness in schizophrenia to a dysfunction of the prefrontal cortex and its connection with cerebellar and thalamic functions necessary to establish common time frames for integrated tasks (Vogeley and Kupke 2007).

Disturbances of Autobiographical (Explicit) Temporality

Disturbances on the higher level of autobiographical temporality are common in many psychopathological conditions. As an example, I will describe what may be called a *fragmentation of the narrative self* in Borderline personality disorder (BPD).

While basic neuropsychological functions of temporality are intact in BPD, the *rapidly changing affects and moods* conspicuous in these patients result in an incoherence of self-related states and self-concepts. Extreme affective oscillations make them feel almost like different people, each defined by a particular mood state. The result is a shifting view of oneself, with sharp discontinuities, rapidly changing roles, goals and relationships, and an underlying feeling of inner emptiness.

There is no sense of personal continuity over time and across situations, no concept of self-development that could be projected into the future, but only an endless repetition of the same affective states, creating a peculiar atemporal mode of existing (Westen and Cohen 1993; Fuchs 2007c). Borderline individuals, as it were, *are* only what they are experiencing at a given moment, in an often intense, and yet empty present; for it lacks the fulfillment which originates from the integration of past experience and anticipated future.

Similarly, the well-known phenomenon of *splitting* (Kernberg 1975; Kernberg et al. 1989) signifies that borderline individuals are unable to integrate positive and negative aspects of the self and others into coherent perceptions. Depending on the present state of affect, the other is either totally good or totally bad, ideal or devalued; the self is either noble or mean, grandiose or corrupt etc. This results in a failure to form a coherent, over-arching self-concept. Though identity disturbances can be found in other types of personality disorders as well, they are typically associated with BPD and present in the majority of cases (60–90%; Wilkinson-Ryan and Westen 2000). Patients describe a painful sense of incoherence and inauthenticity; they feel as if they were only pretending to be what they are, as if they cheated others into believing them. In fact their personality often changes dramatically depending on whom they are with. Even their sexual identity may be unstable and shifting.

The fragmentation of identity is connected to an *incoherence of autobiographical memory* to be found in borderline individuals. They have marked difficulties in recalling specific autobiographical experiences, and often their narrative accounts show large gaps or inconsistencies (Startup et al. 1999; Wilkinson-Ryan and Westen 2000). These difficulties are closely related to the patients' tendency to *dissociate*. Dissociation may be regarded as a failure to integrate perception, affect, memory, and identity into a coherent sense of consciousness and self. There is accumulating evidence that susceptibility to dissociation is, at least in part, the result of traumatic experiences and adverse early environments (Van Ijzendoorn and Schuengel 1996). Dissociated states first manifest themselves when traumatic experiences are initially stored in memory as sensory fragments without a coherent narrative. Dissociation as well as over-general autobiographical recall may serve as a strategy to avert trauma-related distressing emotions. On the other hand, they additionally undermine the coherence of the life narrative. Thus, BPD particularly demonstrates the vulnerable constitution of autobiographical self-coherence.

Disturbances of Intersubjectivity

All mental disorders imply more or less profound disturbances of intersubjectivity, that is, a restricted freedom of behaving and interacting with others in the common life-world. However, the concepts of intersubjectivity currently prevailing in clinical psychology and psychopathology are mainly based on a mentalistic approach that locates the disorder inside the patient. They assume a fundamental strangeness and inaccessibility of the other whose hidden mental states, thoughts or feelings

may only be indirectly inferred from his external bodily behaviour by using a 'Theory of Mind' (ToM), 'mentalizing' or 'mindreading'. On this view, disorders of intersubjectivity e.g. in autism or schizophrenia are derived from a faulty development or functioning of ToM modules. From a phenomenological point of view, however, intersubjectivity is primarily based on a prereflective, immediate relationship of self and other in an emergent bi-personal field. Instead of a theory deficit, autistic and schizophrenic patients rather suffer from a basic disturbance of being-with-others which they try to compensate by a 'morbid rationalism', i.e. precisely by hypothetical constructs and assumptions about the world of the others. Hence, at least two levels of intersubjectivity should be distinguished: (a) *primary intersubjectivity* or 'intercorporeality' (Merleau-Ponty 1960), and (b) *secondary intersubjectivity*, culminating in the achievement of a 'self-other metaperspective' (Laing et al. 1966).

(a) *Primary intersubjectivity* (Trevarthen 1979) develops in the first year of life. Imitation of facial expression starts from birth on, that means, infants are already able to transpose the seen facial expressions of others into their own proprioception and movement (Meltzoff and Moore 1977, 1989). This bodily resonance is supposedly mediated by the mirror neuron system in the premotor cortex (Gallese 2001). Moreover, familiar patterns of interaction and affect attunement are laid down as interactive schemas in implicit memory. Thus, long before the age of 4, the supposed age for acquiring a ToM, the infant already gains a basic understanding of others through common practices. In embodied and empathic interaction, the other is not assumed to be located 'behind' his action, but he enacts and expresses his intentions, and in seeing his expressive movements and actions embedded in their specific context, "… one already sees their meaning. No inference to a hidden set of mental states is necessary" (Gallagher and Zahavi 2008, p. 185). Thus, phenomenology denies the principal divide between the other's mind and body assumed by current theories of social cognition. Bodily behaviour is intentional and meaningful in its context, and as such it is beyond the artificial distinction of internal and external. It constitutes a sphere of primary 'intercorporeality' as the basis for all forms of intersubjectivity.

(b) Around the age of one year, infants increasingly go beyond the mutual resonance of intercorporeality and begin to refer to the common context explicitly, namely by joint attention, gaze-following and pointing. By noticing how others interact with the world, they learn the usage and meaning that objects have for them, and they recognize others' goals and intentions in uncompleted actions (Baldwin and Baird 2001; Meltzoff and Brooks 2001). Thus, the dyadic interaction opens out towards objects in the surrounding field. At the time, triangulating interactions with others emerge – typically, the mother–child dyad is augmented by the father's involvement (Fivaz-Depeursinge and Corboz-Warnerey 1999). Through this '*secondary intersubjectivity*' (Trevarthen and Hubley 1978), infants begin to perceive others as intentional agents whose actions and mutual interactions are meaningful in pragmatic contexts. In the course of cooperative actions, they also experience themselves as being perceived as intentional agents by others, in a common social space that gradually assumes a symbolic structure.

Symbolic interaction is already present in pointing and cooperative action, but reaches its crucial stage in *language.* Verbal narratives then become the presupposition for more sophisticated modes of understanding that develop in the third and fourth year of life (Gallagher and Hutto 2008). By engaging in story-telling practices, children learn to understand others in a meaningful way, to imagine their goals and intentions as underlying a certain course of actions. Narrative competency allows the child to develop the capacities of taking the other's perspective, of pretend playing and role-taking, and, finally, for certain predictive capacities underlying the usual ToM-tasks (e.g. the 'false belief' test). Interpersonal perception in its full senses is based on the ability to freely oscillate between an egocentric, embodied perspective on the one hand, and an allocentric or decentred perspective on the other. This decisive step of human development may be summarized as reaching an 'excentric position' (Plessner 1981) or as adopting a 'self-other metaperspective' (Laing et al. 1966)

In the following, pathologies of both levels of intersubjectivity will be pointed out, taking autism and schizophrenia as paradigm conditions.

Disturbances of Primary Intersubjectivity in Autism

As a paradigm developmental disorder of intersubjectivity, autism has increasingly become a topic of discussion in phenomenology as well as in cognitive neuroscience. The present psychopathology of autism is still dominated by a cognitive and modular approach, assuming a faulty development of ToM-modules that leads to a disturbed capacity to attribute mental states to others (Baron-Cohen 1995; Frith 1989). In recent years, however, criticism has been raised by phenomenological psychiatrists and philosophers (Hobson 1993, 2002; Gallagher 2004b), arguing that the deficit already involves failures of early interaction and interaffectivity. This is supported by the fact that many autistic symptoms such as lack of emotional contact, anxiety or agitation are already present in the first years of life, i.e. long before the supposed age to acquire a ToM. Moreover, between 15% and 60% of autistic individuals are able to pass false belief tests successfully, pointing out that the disorder can hardly be due only to a lack of ToM (Reed and Paterson 1990). From a phenomenological approach, autism should rather be conceived as a *disorder of primary or embodied intersubjectivity*. This includes disturbances in (a) sensory-motor integration, (b) imitation and affect attunement, and (c) holistic perception.

(a) There is evidence that autistic children show a variety of basic sensory-motor abnormalities on the neurological level (Vilensky et al. 1981; Mari et al. 2003). In studies of videotapes such abnormalities could already be found in the first year of life in children who were later diagnosed as autistic (Teitelbaum et al. 1998), e.g. problems in righting, sitting, crawling and walking, or abnormal motor patterns. This points to a *deficient integration* of visual, kinesthetic, vestibular and tactile sensations into a common experiential space. Infant research has shown that early dyadic interactions are particularly based on the integration of sensory, motor and affective experience (Stern 1985). In other words, there is a close connection

between the *'sensus communis'* (intermodal integration) and social attunement or *'common sense'*. Hence, faulty intermodal integration may significantly interfere with the development of embodied social cognition in autistic children.

(b) A particular aspect of disturbed integration concerns the sensory-motor feed-back loops involved in *imitation*. Based on the mirror neuron system and shared self-other representations of movement (Decety and Sommerville 2003), imitation serves as a major instrument for early social cognition (Meltzoff 2002). The literature shows a consistent finding that people with autism do not readily imitate the actions of others (Smith and Bryson 1994; Hobson and Lee 1999). There is also increasing evidence for a mirror neuron dysfunction in autism spectrum disorders (Oberman et al. 2005; Dapretto et al. 2006). Problems with imitation might then lead to a cascade of impairments in early intercorporeality, affect attunement, joint attention, pretend play and, finally, acquisition of a ToM.

(c) Moreover, autistic children show problems in establishing *perceptual and situational coherence:* They focus on single parts or elements rather than perceiving the Gestalt of objects, and they tend to treat things as decontextualized, thus missing their particular meaning provided by the situation as whole (Frith 1989; Happé 1995). While this failure of holistic cognition may have some positive effects such as remembering unrelated or non-sensical items, it significantly interferes with the development of social understanding. Thus, affect attunement is crucially based on perceiving emotional cues (gestures, facial expression, vocal intonations) as embedded in recurrent situations. Even more, secondary intersubjectivity depends on learning how to relate gestures and actions of others to the context in order to grasp their intentions. Correspondingly, eye tracking studies have shown that autistic children focus on inanimate and irrelevant details of interactive situations while missing the relevant social cues (Klin et al. 2003).

Although the question of reciprocal interaction between these different mechanisms is as yet far from being solved, it seems most likely that they converge to a fundamental disturbance of embodied social cognition very early in life. This disturbance is then bound to compromise the later stages of intersubjectivity; for these are not based on ToM modules that develop separately, but rather on a primary 'sensus communis' or 'social sense' that is subsequently extended by reference to the common context, by triangulating interactions and by understanding others as intentional agents like oneself. In sum, what autistic children lack is not a theoretical concept of others' minds; on the contrary, ToM-like strategies of mentalizing and inferring from social cues are rather employed by high-functioning autistic individuals as a compensation for the lacking capacities of primary and secondary intersubjectivity (Zahavi and Parnas 2003).

Disturbances of Secondary Intersubjectivity in Schizophrenia

According to currently dominant theories, schizophrenia should likewise be understood as involving some incapacity for meta-awareness, self-monitoring and theory

of mind. Frith (1992) has proposed that schizophrenia can be explained by impaired metarepresentation: Problems with monitoring one's own intentions to think or act result in symptoms such as thought-insertion or delusions of alien control. Moreover, the inability to correctly infer the mental states of others gives rise to paranoid delusions. A number of experimental studies have shown that patients with schizophrenia perform badly in theory of mind tasks (see Lee et al. 2004 for a review). However, studies on real life interactions could not confirm these results – in normal conversations even delusional patients showed intact theory of mind skills (Walston et al. 2000; McCabe et al. 2004). Obviously, the interpretation of the results depends on how one conceives the role of narrative and context versus abstract mentalizing abilities in understanding others (Gallagher and Hutto 2008).

In contrast to meta-representational approaches to schizophrenia, phenomenological psychopathology, as we have seen, emphasizes disturbances of basic self-awareness and attunement to the social world. Pathologies of the prereflective, embodied self necessarily also impair the patient's ability to interact with others. What is lacking in schizophrenic autism, then, is not explicit knowledge, inferential or ToM abilities, but rather an implicit understanding of the 'rules of the game', a sense of proportion for what is appropriate, likely and relevant in the social context (Parnas et al. 2002). However, the disturbance of basic self-awareness does not only affect primary intersubjectivity, but also the higher level of self-other distinction or self-demarcation, resulting in phenomena termed *transitivism* by Bleuler (1911):

> When I look at somebody my own personality is in danger. I am undergoing a transformation and my self is beginning to disappear (Chapman 1966).

> The others' gazes get penetrating, and it is as if there was a consciousness of my person emerging around me … they can read in me like in a book. Then I don't know who I am any more (Fuchs 2000, 172).

Such reports show that 'being conscious of another consciousness' may threaten the schizophrenic patient with a loss of his self. How could this be explained? – In current neurocognitive accounts, the sense of self is regarded as being generated by inferential self-monitoring processes. Corresponding explanation of symptoms such as transitivism, thought insertion, acoustic hallucinations or passivity experiences rely on the concept of *shared representations*, i.e. overlapping neuronal representations for the execution of an action and for the observation of the same action in others (Decety and Sommerville 2003). A hypothetical failure of the action attribution system (neuronal "who" system, Georgieff and Jeannerod 1998) then leads to self-other confusion and delusional misattribution.

However, such modular explanations miss the basic disturbance of self-awareness that precedes the acute psychotic symptoms often by years. From a phenomenological perspective, the self-other distinction is automatically constituted in every experience as an aspect of non-reflective self-awareness (Parnas 2003). If this primary sense of self or ipseity is disturbed, then taking a self-other metaperspective will become precarious. In grasping the other's perspective, the patients are no more able to maintain their own embodied center:

> A young man was frequently confused in a conversation, being unable to distinguish between himself and his interlocutor. He tended to lose the sense of whose thoughts originated

in whom, and felt 'as if' the interlocutor somehow 'invaded' him, an experience that shattered his identity and was intensely anxiety-provoking. When walking on the street, he scrupulously avoided glancing at his mirror image in the windowpanes of the shops, because he felt uncertain on which side he actually was (Parnas 2003, p. 232).

Interpersonal perception, as we have seen, implies a continuous oscillation between the central, embodied perspective on the one hand, and the decentred or virtual perspective on the other. The same applies for perceiving oneself in the mirror. It is this dialectical tension of the "excentric position" that the schizophrenic patient cannot keep up any more. The perspectives of self and other are confused instead of being integrated. This short-circuit of perspectives may also lead to the experience of thought-broadcasting: All the patient's thoughts are known to others; there is no difference between his mental life and that of others any more.[10] Thus, he is entangled in a disembodied, self-referential and delusional view from the outside (Fuchs 2002a, 2005a). It seems most likely that this short-circuit should also be mirrored in faulty activity of the neural systems representing self- and other-generated actions. However, the phenomenological approach would emphasize the precedent weakness on the level of basic self-awareness.

It is also for this reason that schizophrenia manifests itself often in situations of social exposure and emotional disclosure, when the affirmation of one's own self against the perspective of the others is at stake: e.g. when leaving the parents' home, starting an intimate relationship or entering working life. In such situations, the patient may lose his embodied perspective and start to feel observed, persecuted and permeated from all sides. Thus we find again what may be called a *disembodiment,* caused by a loss of self in the dialectical process of intersubjective perception. There are two psychopathological outcomes of this alteration. First, schizophrenic *autism* may be caused by a withdrawal from the threatening intersubjective sphere. Second, *delusions* constitute a "locked reality" that protects the deluded person from being questioned and overwhelmed by the others' intentionality. In the last analysis, intersubjectivity can only be maintained at the price of a severe restriction of the patient's life world.

Conclusion

From a phenomenological point of view, mental illness is not something merely "mental" but manifests itself in dimensions such as self-awareness, embodiment, temporality and intersubjectivity, or in short, in an alteration of the patient's overall being-in-the-world. I have argued that these dimensions may each be characterized by a duality of levels: a basic, implicit level of primary experience, and an extended, explicit level of conscious relation to oneself and to others. On both levels as well as in their relation to each other disturbances may arise, and it is of essential diagnostic and therapeutic relevance to adequately attribute the actual illness to those

---

[10] A possible failure of the neuronal "who" system may then serve as a pathway, but not as the primary cause of the self-other confusion.

levels.[11] The symptoms that manifest themselves on the explicit and verbal level are not always suited to reflect the primary disturbances of experience. In contrast, phenomenology is capable of grasping the level of prereflective experience which is concerned particularly in psychotic disorders and whose alteration may secondarily generate the productive symptoms.

On the other hand, the patient's relation to himself is an essential component of psychopathology as well. For to a certain extent, mental illness always means a self-alienation. Something within me opposes me, escapes my control and dominates me, while I am in vain trying to regain my autonomy. Therefore the subjective side of the illness does not only consist in a secondary reaction to some basic physiological dysfunction as is the case in somatic diseases; it is inseparable from the illness itself. Thus, it may include, e.g. in depression, negative self-evaluations and thought patterns which, on their part, act as self-fulfilling prophecies and thus aggravate the depression. For disorders such as social phobia, hypochondriasis or anorexia it is even more obvious that they are grounded in the patient's self-relation. Finally, schizophrenia is not just a bundle of dysfunctions but involves a profound transformation of the self that includes particular forms of intentionality, struggling and coping with one's illness (Sass 2007). In any case, subjectivity as a self-relation implying the continuous necessity of taking a stance towards one's own state does not permit us to regard mental disorders as mere biological dysfunctions. This is also the presupposition for every psychotherapeutic intervention that addresses the patient as a self-relating subject.

Phenomenology further challenges core assumptions of present psychiatry by overcoming the narrow conception of the patient as an enclosed individual with a clearly defined brain dysfunction and by recognizing the ways in which the disorder in question is being shaped by the patient's intersubjective, socio-historical situation. Independent of its etiology, mental illness is always a disturbance of the person's relations to others. It is accompanied by various restrictions of one's freedom to respond to the social environment in a flexible and autonomous way. Thus, mental disorders may also be conceived as disturbances of *responsivity* (Fuchs 2007a). A person's social capacities are either inhibited by her illness or have primarily not developed in such a way as to enable her to regulate her relationships in a satisfying way. Therefore, a considerable part of psychopathology may not be assessed in the individual patient, let alone his brain, but only in his interactions with others. From this it follows that the simple bottom-up explanation of mental disorders as products of genetic or neurophysiological determinants is inadequate to the causal complexity involved. From an ecological or systemic perspective, the disorder should rather be regarded as the product of a *circular causality* of neurophysiological, subjective, environmental and social influences continuously interacting with each other.[12]

---

[11]To take an example: ignoring the difference between primary and extended self has led to an industry of neuroimaging studies that claim to investigate the first-person perspective while only being able to access higher order self-referential judgments (Mishara 2007).

[12]Circular models involving negative feedback loops of primary symptoms, emotions, cognitions and social interactions have been developed e.g. for depressive, anxiety or Borderline Personality disorders (Linehan 1993; Grawe 2002).

Phenomenology regards the person as inseparable from their being-in-the-world and being-with-others. This is in correspondence with a systemic or ecological view of the brain as an organ that is embedded in, and continuously shaped by environmental relations. The increasing convergence of phenomenology and embodied or enactive cognitive neuroscience also applies for phenomenological psychopathology and systemic accounts of mental illness. The potentialities held out by a close cooperation between these approaches are just being discovered.

References

Andreason NC, Paradiso S, O'Leary DS (1998) "Cognitive dysmetria" as an integrative theory of schizophrenia: a dysfunction in cortical-subcortical-cerebellar circuitry? Schizophr Bull 24:203–218

Baldwin DA, Baird JA (2001) Discerning intentions in dynamic human action. Trends Cogn Sci 5:171–178

Baron-Cohen S (1995) Mindblindness: an essay on autism and theory of mind. MIT Press, Cambridge, MA

Bech P (1975) Depression: influence on time estimation and time experience. Acta Psychiatrica Scandinavia 51:42–50

Benninghoven D, Raykowski L, Solzbacher S, Kunzendorf S, Jantschek G (2007) Body images of patients with anorexia nervosa, bulimia nervosa and female control subjects: a comparison with male ideals of female attractiveness. Body Image 4:51–59

Blankenburg W (1971) Der Verlust der natürlichen Selbstverständlichkeit (the loss of natural self-evidence). Springer, Berlin

Blankenburg W (2001) First steps toward a 'psychopathology of common sense' (transl: Mishara A). Philos Psychiatry Psychol 8:303–315

Bleuler E (1911) Dementia Praecox oder Gruppe der Schizophrenien. In: Aschaffenburg G (ed) Handbuch der Psychiatrie. Franz Deuticke, Leipzig

Braus DF, Teckhoff A, Tost H (2001) Alteration of temporal order threshold in schizophrenia. Biol Psychiatry 49:S121–S122

Buhlmann U, Cook LM, Fama JM, Wilhelm S (2007) Perceived teasing experiences in body dysmorphic disorder. Body Image 4:381–385

Damasio AR (1999) The feeling of what happens. Harcourt, San Diego, CA

Dapretto M, Davies MS, Pfeifer JH, Scott AA, Sigman M, Bookheimer SY, Jacoboni M (2006) Understanding emotions in others: mirror neuron dysfunction in children with autims spectrum. Nature Neuroscience online (www.nature.com/natureneuroscience)

Decety J, Sommerville JA (2003) Shared representations between self and other: a social cognitive neuroscience view. Trends Cogn Sci 7:527–533

Dignon A, Beardsmore A, Spain S, Kuan A (2006) Why I wont eat: patient testimony from 15 anorexics concerning the causes of their disorder. J Health Psychol 11:942–956

Enoch MD, Trethowan WH (1991) Uncommon psychiatric syndromes. John Wright, Bristol

Fivaz-Depeursinge E, Corboz-Warnerey A (1999) The primary triangle. Basic Books, New York

Frith CD (1992) The cognitive neuropsychology of schizophrenia. Earlbaum, Hillsdale

Frith CD (2004) Schizophrenia and theory of mind. Psychol Med 34:385–389

Frith U (1989) Autism: explaining the enigma. Basil Blackwell, Oxford

Fuchs T (1995) Coenästhesie. Zur Geschichte des Gemeingefühls. Z Klin Psychol Psychopath Psychother 43:103–112

Fuchs T (2000) Psychopathologie von Leib und Raum. Phänomenologisch-empirische Untersuchungen zu depressiven und paranoiden Erkrankungen. Steinkopff, Darmstadt

Fuchs T (2001a) Melancholia as a desynchronization. Towards a psychopathology of interpersonal time. Psychopathology 34:179–186

Fuchs T (2001b) The tacit dimension. Commentary to W. Blankenburg's "Steps towards a psychopathology of common sense". Philos Psychiatry Psychol 8:323–326

Fuchs T (2002a) The challenge of neuroscience: psychiatry and phenomenology today. Psychopathology 35:319–326

Fuchs T (2002b) The phenomenology of shame, guilt and the body in body dysmorphic disorder and depression. J. Phenomenol Psychol 33:223–243

Fuchs T (2002c) Mind, meaning and the brain. Philos Psychiatry Psychol 9:261–264

Fuchs T (2005a) Corporealized and disembodied minds. A phenomenological view of the body in melancholia and schizophrenia. Philos Psychiatry Psychol 12:95–107

Fuchs T (2005b) Implicit and explicit temporality. Philos Psychiatry Psychol 12:195–198

Fuchs T (2005c) Delusional mood and delusional perception – a phenomenological analysis. Psychopathology 38:133–139

Fuchs T (2007a) Psychotherapy of the lived space. A phenomenological and ecological concept. Am J Psychotherapy 61:432–439

Fuchs T (2007b) The temporal structure of intentionality and its disturbance in schizophrenia. Psychopathology 40:229–235

Fuchs T (2007c) Fragmented selves. Temporality and identity in borderline personality disorder. Psychopathology 40:379–387

Fuster JM (1997) The prefrontal cortex. Anatomy, physiology, and neuropsychology of the frontal lobe, 3rd edn. Lippincott-Raven, Philadelphia/New York

Fuster JM (2003) Cortex and mind – unifying cognition. Oxford University Press, Oxford/UK

Gadamer H-G (1989) Truth and method. 2nd revised edn (trans: Weinsheimer J, Marshall DG). Crossroad, New York

Gallagher S (2000a) Philosophical conceptions of the self: implications for cognitive science. Trends Cogn Sci 4:14–21

Gallagher S (2000b) Self-reference and schizophrenia: a cognitive model of immunity to error through misidentification. In: Zahavi D (ed) Exploring the self: philosophical and psychopathological perspectives on self-experience. John Benjamins, Amsterdam, Philadelphia, pp 203–239

Gallagher S (2004a) Neurocognitive models of schizophrenia: a neurophenomenological critique. Psychopathology 37:38–19

Gallagher S (2004b) Understanding interpersonal problems in autism: interaction theory as an alternative to theory of mind. Philos Psychiatry Psychology 11:199–217

Gallagher S (2005) How the body shapes the mind. Clarendon, New York

Gallagher S, Hutto D (2008) Understanding others through primary interaction and narrative practice. In: Zlatev, Racine, Sinha, Itkonen (eds) The shared mind: perspectives on intersubjectivity. John Benjamins, Amsterdam, pp 17–38

Gallagher S, Vaever M (2004) Disorders of embodiment. In: Radden J (ed) The philosophy of psychiatry. A companionpp. Oxford University Press, Oxford, pp 118–132

Gallagher S, Zahavi D (2005) Phenomenological approaches to self-consciousness. Stanford encyclopedia of philosophy (http://plato.stanford.edu)

Gallagher S, Zahavi D (2008) The phenomenological mind. An introduction to philosophy of mind and cognitive science. Routledge, London/New York

Gallese V (2001) The 'shared manifold' hypothesis. From mirror neurons to empathy. J Conscious Stud 8:33–50

Georgieff N, Jeannerod M (1998) Beyond consciousness of external events: A who system for consciousness of action and self-consciousness. Conscious Cogn 7:465–477

Grawe K (2002) Psychological psychotherapy. Hogrefe, Seattle, WA

Gross G, Huber G, Klosterkötter J, Linz M (1987) Bonner Skala für die Beurteilung von Basissymptomen. Springer, Berlin

Happé F (1995) Autism: an introduction to psychological theory. Harvard University Press, Cambridge

Harrington DL, Haaland KY, Knight RT (1998) Cortical networks underlying mechanisms of time perception. J Neurosci 18:1095–1095

Henry M (2000) Incarnation. Une philosophie de la chair. Seuil, Paris

Hobson RP (1993) Autism and the development of mind. Lawrence Erlbaum, Hillsdale, NJ

Hobson P (2002) The cradle of thought. Macmillan, London

Hobson RP, Lee A (1999) Imitation and identification in autism. J Child Psychol Psychiat 40:649–659

Husserl E (1966) Analysen zur passiven Synthesis. Aus Vorlesungs-und Forschungsmanuskripten. Martinus Nijhoff, Husserliana XI, Den Haag, pp 1918–1926

Husserl E (1973) Zur Phänomenologie der Intersubjektivität I, *Husserliana XIII*. Martinus Nijhoff, Den Haag

Husserl E (1991) Lectures on the phenomenology of the consciousness of internal time (1893–1917) (trans: Brough JB). Kluwer Academic, Dordrecht, Boston

Jacobson K (2007) The interpersonal expression of human spatiality. A phenomenological interpretation of Anorexia nervosa. Chiasmi Int 8:157–174

Jirsa R, Libiger J, Mohr P, Radil T, Indra M (1996) Rhythmic finger-tapping task and fast segmentation of neural processing in schizophrenics. Biol Psychiatry 40:1301–1304

Kernberg OF (1975) Borderline conditions and pathological narcissism. Jason Aronson, New York

Kernberg OF, Selzer MA, Koenigsberg HW, Carr AC, Appelbaum A (1989) Psychodynamic psychotherapy of borderline patients. Basic Books, New York

Kitamura T, Kumar R (1982) Time passes slowly for patients with depressive state. Acta Psychiatrica Scandinavia 65:415–420

Klin A, Jones W, Schultz R, Volkmar F (2003) The enactive mind, or from actions to cognition: lessons from autism. Phil Trans R Soc Lond 358:345–360

Kraus A (2002) Melancholie: eine Art von Depersonalisation? In: Fuchs T, Mundt C (eds) Affekt und affektive Stoerungen. Schoeningh, Paderborn, pp 169–186

Laing RD, Phillipson H, Lee AR (1966) Interpersonal perception. Tavistock, London

Leder D (1992) The absent body. University of Chikago Press, Chikago

Lee KH, Farrow TFD, Spence SA, Woodruff PWR (2004) Social cognition, brain networks and schizophrenia. Psychol Med 34:391–400

Linehan M (1993) Cognitive-behavioral treatment of borderline personality disorder. Guilford, New York

Manoach DS (2003) Prefrontal cortex dysfunction during working memory performance in schizophrenia: reconciling discrepant findings. Schizophr Res 60:285–298

Marenco S, Weinberger DR (2000) The neurodevelopmental hypothesis of schizophrenia: following a trail of evidence from cradle to grave. Devel Psychopathol 12:501–527

Mari M, Castiello U, Marks D, Marraffa C, Prior M (2003) The reach-to-grasp movement in children with autism spectrum disorder. Philos Trans R Soc: Biol Sci 358:393–403

McCabe R, Leudar I, Antaki C (2004) Do people with schizophrenia display theory of mind deficits in clinical interactions? Psychol Med 34:401–412

Mead GH (1924) Mind, self and society. University of Chikago Press, Chicago, IL

Meltzoff AN, Brooks R (2001) 'Like me' as a building block for understanding other minds: bodily acts, attention, and intention. In: Malle BF, Moses LJ, Baldwin DA (Hrsg.) Intentions and intentionality: foundations of social cognition. MIT, Cambridge, MA, S. 171–191

Meltzoff AN, Moore MK (1977) Imitation of facial and manual gestures by human neonates. Science 198:74–78

Meltzoff A, Moore MK (1989) Imitation in newborn infants: exploring the range of gestures imitated and the underlying mechanisms. Dev Psychol 25:954–962

Meltzoff AN (2002) Elements of a developmental theory of imitation. In: Meltzoff AN, Prinz W (eds) The imitative mind: development, evolution, and brain bases. Cambridge University Press, Cambridge, pp 19–41

Merleau-Ponty M (1945) Phénoménologie de la perception. Éditions Gallimard; English translation, Paris; Smith C (1962) Phenomenology of perception. Routledge and Kegan Paul, London

Merleau-Ponty M (1960) Le philosophe et son ombre. In: Éditions Gallimard. Signes, Paris

Mezey AG, Knight EJ (1965) Time sense in hypomanic illness. Arch Gen Psychiatry 12:184–186

Mishara AL, Parnas J, Naudin J (1998) Forging the links between phenomenology, cognitive neuroscience, and psychopathology: the emergence of a new discipline. Curr Opin Psychiatry 11:567–573

Mishara AL (2007) Missing links in phenomenological clinical neuroscience: why we are not there yet. Curr Opin Psychiatry 20:559–569

Münzel K, Gendner G, Steinberg R, Raith L (1988) Time estimation of depressive patients: the influence of the interval content. Eur Arch Psychiat Neurol Sci 237:171–178

Mundt C, Richter P, van Hees H, Stumpf T (1998) Zeiterleben und Zeitschaetzung depressiver Patienten. Nervenarzt 69:38–45

Neisser U (1988) Five kinds of self-knowledge. Philos Psychol 1:35–39

Parnas J, Sass LA (2001a) Solipsism, self, and schizophrenic delusions. Philos Psychiatry Psychol 8:101–120

Oberman LM, Hubbard EM, McCleery JP, Altschuler EL, Ramachandran VS, Pineda JA (2005) EEG evidence for mirror neuron dysfunction in autism spectrum disorders. Cogn Brain Res 24:190–198

Parnas J (2003) Self and schizophrenia: a phenomenological perspective. In: Kircher T, David A (eds) The Self in Neuroscience and Psychiatry. Cambridge University Press, Cambridge, pp 217–241

Parnas J, Bovet P (1995) Research in psychopathology: epistemological issues. Compr Psychiatry 36:167–181

Parnas J, Bovet P, Innocenti G (1996) Schizophrenic trait features, binding and cortico-cortical connectivity: a neurodevelopmental pathogenetic hypothesis. Neurol Psychiatry Brain Res 6:97–106

Parnas J, Bovet P, Zahavi D (2002) Schizophrenic autism: clinical phenomenology and pathogenetic implications. World Psychiatry 1:131–136

Parnas J, Moeller P, Kircher T, Thalbitzer J, Jannson L, Handest P, Zahavi D (2005a) EASE: examination of anomalous self-experience. Psychopathology 38:236–258

Parnas J, Sass LA (2001b) Solipsism, self, and schizophrenic delusions. Philos Psychiatry Psychol 8:101–120

Parnas J, Moeller P, Kircher T, Thalbitzer J, Jannson L, Handest P, Zahavi D (2005b) EASE: examination of anomalous self-experience. Psychopathology 38:236–258

Pinto A, Phillips KA (2005) Social anxiety in body dysmorphic disorder. Body Image 2:401–405

Phillips KA (2000) Personality disorders and traits in patients with body dysmorphic disorder. Compr Psychiatry 41:229–236

Phillips KA, Pinto A, Jain S (2004) Self-esteem in body dysmorphic disorder. Body Image 1:385–390

Philipps J (2003) Psychopathology and the narrative self. Philos Psychiatry Psychol 10:313–328

Plessner H (1981) Die Stufen des Organischen und der Mensch. Gesammelte Schriften IV. Suhrkamp, Frankfurt am Main

Polanyi M (1967) The tacit dimension. Anchor Books, Garden City, New York

Reed T, Paterson C (1990) A comparative study of autistic subjects' performance at two levels of visual and cognitive perspective taking. J Autism Dev Disord 29:555–568

Rochat P (2004) The emergence of self-awareness as co-awareness in early child development. In: Zahavi D, Grünbaum T, Parnas J (eds) The structure and development of self-consciousness. John Benjamins, Philadelphia, PA

Sartre J-P (1943) L'Etre et le néant. Tel Gallimard, Paris. English translation (1956) Barnes HE Being and nothingness. Philosophical Library, New York

Sass LA (1992) Madness and modernism. Insanity in the light of modern art, literature, and thought. Basic Books, New York

Sass LA (2000) Schizophrenia, self-experience, and so-called negative symptoms. In: Zahavi D (ed) Exploring the self: philosophical and psychopathological perspectives on self-experience. John Benjamins, Amsterdam, pp 149–182

Sass LA, Parnas J (2003) Schizophrenia, consciousness, and the self. Schizophr Bull 29:427–444

Sass LA (2007) 'Schizophrenic person' or 'person with schizophrenia'? Theory Psychol 17:395–420

Seidler GH (2000) In others' eyes: an analysis of shame. International Universities Press, Boston

Skrzypek S, Wehmeier PM, Remschmidt H (2001) Body image assessment using body size estimation in recent studies on anorexia nervosa. A brief review. Eur Child Adolesc Psychiatry 10:215–21

Smith IM, Bryson SE (1994) Imitation and action in autism: a critical review. Psychol Bull 116:259–73

Spiegelberg H (1972) Phenomenology in psychology and psychiatry. a historical introduction. Northwestern University Press, Evanston

Stanghellini G (2004) Disembodied spirits and deanimatied bodies: the psychopathology of common sense. Oxford University Press, Oxford

Stanghellini G (2007) The grammar of the psychiatric interview. A plea for the second-person mode of understanding. Psychopathology 40:69–74

Startup M, Swales M, Williams JMG, Jones RSP (1999) Autobiographical memory and dissociation in borderline personality disorder. Psychol Med 29:1397–1404

Stern D (1985) The interpersonal world of the infant. Basic Books, New York

Tauscher-Wisniewski S (1999) Cognitive processing speed slows before schizophrenia, Poster session, Society of Biological Psychiatry; reported by Moon MA. Clin Psychiatry News 27(7):1

Taylor C (1989) Sources of the self: the making of the modern identity. Cambridge University Press, Cambridge

Teitelbaum P, Teitelbaum O, Nye J, Fryman J, Maurer RG (1998) Movement analysis in infancy may be useful for early diagnosis of autism. Proc Nat Acad Sci USA 95:13982–13987

Tellenbach H (1980) Melancholy. History of the problem, endogeneity, typology, pathogenesis, clinical considerations. Duquesne University Press, Pittsburgh

Trevarthen C, Hubley P (1978) Secondary intersubjectivity: confidence, confiding and acts of meaning in the first year. In: Lock A (ed) Action, gesture and symbol: the emergence of language. Academic, London, pp 183–229

Trevarthen CB (1979) Communication and cooperation in early infancy: a description of primary intersubjectivity. In: Bullowa M (ed) Before speech. Cambridge University Press, Cambridge

Uhlhaas PJ, Mishara AL (2007) Perceptual anomalies in schizophrenia: integrating phenomenology and cognitive neuroscience. Schizophr Bull 33:142–156

Van Ijzendoorn MH, Schuengel C (1996) The measurement of dissociation in normal and clinical populations: meta-analytic validation of the Dissociative Experience Scale (DES). Clin Psychol Rev 16:365–382

Vilensky JA, Damasio AR, Maurer RG (1981) Gait disturbances in patients with autistic behavior: a preliminary study. Arch Neurol 38:464–469

Vogeley K, Bussfeld P, Newen A, Herrmann S, Happé F, Falkai P et al (2001) Mind reading: neural mechanisms of theory of mind and self-perspective. Neuroimage 14:170–181

Vogeley K, Kupke C (2007) Disturbances of time consciousness from a phenomenological and a neuroscientific perspective. Schizophr Bull 33:157–165

Vogeley K, Kurthen M, Falkai P, Maier W (1999) The human self construct and prefrontal cortex in schizophrenia. The Association for the scientific study of consciousness: electronic seminar (http://www.phil.vt.edu/assc/esem.html)

Walston F, Blennerhassett RC, Charlton B (2000) 'Theory of mind', persecutory delusions and the somatic marker mechanism. Cogn Neuropsychiatry 5:161–174

Walter G, Sines J, Meyer C, Foster E, Skelton A (2007) Narcissism and narcissistic defenses in the eating disorders. Int J Eat Disord 40:143–148

Westen D, Cohen RP (1993) The self in borderline personality disorder: a psychodynamic perspective. In: Segal ZS, Blatt SJ (eds) The self in emotional distress. Cognitive and psychodynamic perspectives. New York, Guilford, pp 334–360

Wiggins OP, Schwartz MA (2007) Schizophrenia: a phenomenological-anthropological approach. In: Chung MD, Fulford KWM, Graham G (eds) Reconceiving schizophrenia. Oxford University Press, Oxford, pp 113–127

Wiggins OP, Schwartz MA, Northoff G (1990) Toward a Husserlian phenomenology of the initial stages of schizophrenia. In: Spitzer M, Maher BA (eds) Philosophy and psychopathology. Springer, Berlin Heidelberg, New York, pp 21–34

Wilkinson-Ryan T, Westen D (2000) Identity disturbance in borderline personality disorder: an empirical investigation. Am J Psychiatry 157:528–541

Zahavi D (1999) Self-awareness and alterity. A phenomenological investigation. Northwestern University Press, Evanstone

Zahavi D (2005) Subjectivity and selfhood: investigating the first-person perspective. MIT, Cambridge, MA

Zahavi D, Parnas J (2003) Conceptual problems in infantile autism research: why cognitive science needs phenomenology. J Conscious Stud 10:53–71

Delusional Atmosphere and Delusional Belief

Matthew Ratcliffe

Introduction

Delusions are usually taken to be mistaken beliefs, which are inferred from experience and are highly resistant to change, despite overwhelming evidence against them (e.g. DSM-IV-TR, p. 821). However, there is some controversy as to whether the term 'delusion' refers to a distinctive kind of mental state or to a range of different psychological predicaments, and no definition of delusion is uncontroversial. As David (1999, p. 17) remarks, "despite the façade created by psychiatric textbooks, there is no acceptable (rather than accepted) definition of a delusion".[1] Furthermore, it is not always clear what is meant by the claim that delusions are *beliefs* (see, for example, Bayne and Pacherie 2005; Bortolotti 2005).

Despite such concerns, much the same strategy has been employed to explain a range of delusions. Delusions, it is maintained, are the product of two different impairments. First of all, there is an anomalous perception. This is then fed into defective reasoning processes, which generate a delusional belief. It is generally assumed that the 'belief' in question is a propositional attitude and that the delusion takes the form 'X believes that p' where p is a false proposition that any rational being with access to the same information ought to reject. Approaches like this tend to presuppose a rather impoverished conception of experience. It is construed as a kind of input system that presents the subject with assorted perceptual contents, which are then fed into belief-forming processes.

In this chapter, I will suggest that we can acquire a much richer understanding of the experiential changes involved in delusions by drawing on the resources of phenomenology. Phenomenologists emphasise that experience is not just a matter of having various perceptual contents. Amongst other things, it also incorporates a background sense of belonging to the world. Most and possibly all delusions involve changes to this background, *existential* shifts that result in the deluded

M. Ratcliffe
Department of Philosophy, Durham University, UK

[1] See also Fulford (1994) and Broome (2004) for the view that the term 'delusion' can refer to quite different kinds of mental state.

person no longer *finding herself in the world* in the way she previously did. Some delusions, I will propose, are not propositional attitudes at all but *expressions* of altered existential orientations. These expressions involve at least some degree of *interpretation*, given that the same experience can be expressed in different ways. Other delusions, although not so closely associated with existential changes, are still *symptomatic* of those changes and need to be understood in terms of them.

In what follows, I will draw on three interconnected themes in phenomenology (which feature in the work of several phenomenologists) and show how they can be applied so as to cast light on the nature of delusional experience:

1. Experience of one's body and experience of the world are inseparable. Changed bodily feeling can also be a way in which things outside of the body are experienced.
2. In addition to experiencing the actual, we also experience the possible.
3. Nothingness or absence can be a conspicuous feature of an experience.

My primary focus will be the Capgras and Cotard delusions, and I will suggest that both delusions should be understood as expressions of disturbances in the existential structure of experience, rather than as beliefs that arise due to anomalous perceptual contents combined with defective reasoning. I will concede that this account will not apply to all cases of delusion. Even so, I will suggest that all or almost all delusions owe at least something to existential changes.

The Capgras and Cotard Delusions

The Capgras delusion is characterised by the belief that one or more familiars (usually a spouse and/or other family members) have been replaced by impostors. These impostors might take on a more specific form, such as robots or aliens. A core 'belief' common to all instances of the Cotard delusion is harder to pin down. Young and Leafhead (1996) focus on the claim that one is dead. However, they also acknowledge that patients describe their predicaments in a range of different ways – as damned, disembodied, non-existent, consisting only of a voice, and so on.

Many recent discussions of these delusions draw upon a model first proposed by Ellis and Young (1990). According to this model, the Capgras delusion is a 'mirror image' of prosopagnosia. Patients with the latter condition are unable to explicitly recognise familiar faces. However, they do show a pronounced affective response to these faces (which can be detected by measuring skin conductance response). Ellis and Young relate this finding to the view that there are two different neural pathways involved in visual recognition, a 'ventral route' that facilitates overt recognition and a 'dorsal route' that generates a covert, affective response. They hypothesise that the ventral route is affected in prosopagnosia whereas the dorsal route is damaged in the Capgras delusion. Hence Capgras patients have intact overt recognition but the associated affective response is diminished or absent, resulting in an experience of familiar faces as somehow different, strangely unfamiliar.

The same model has also been applied to the Cotard delusion. For instance, Stone and Young (1997) suggest that both delusions involve an absence of affective response to familiar faces and a consequent experience of strangeness and unfamiliarity. They concede that the experience alone is not a sufficient cause for either delusional belief. No rational person would adopt the inflexible belief that she is dead or that her spouse has been replaced by an impostor on the basis that people look strangely unfamiliar. They therefore propose that both delusions also involve impaired reasoning processes, which get subjects from the anomalous experience to the delusional belief. The differing delusional contents are symptomatic of different reasoning biases. Capgras patients are suspicious and prone to persecutory feelings, disposing them to misattribute their unpleasant experiences to changes in the external world. Cotard patients, in contrast, are extremely depressed and so tend to explain the experience in terms of changes to themselves. In addition to this, in both cases there is a bias towards observational adequacy over belief conservatism. In other words, patients incautiously accept explanations that cohere with their experiences, despite conflict between these explanations and the largely coherent framework of beliefs that they built up before they became delusional.

Many subsequent accounts of these and other delusions have adopted the same general two-stage approach, appealing to altered experience followed by a reasoning process of some kind. However, there is disagreement over the nature of the reasoning bias and even over whether there is a bias. For example, Davies et al. (2001) suggest that the Capgras delusion is not an inference from an experience. Rather, the content of the delusion is part of the experience. This content is then unthinkingly accepted in the form of a belief, due to a failure to inhibit the default strategy of assuming that perceptual contents are veridical. And Maher (1999) argues that delusions are primarily a matter of experience, rather than reasoning, and that they can be explained in terms of changed perception in conjunction with completely *normal* reasoning processes.

If the Capgras and Cotard delusions are to be explained in terms of altered experience alone, one option is to maintain that reasoning strategies in the normal population are diverse, allowing very different delusions to arise from the same experiential content. Alternatively, it could be conceded that, contrary to Stone and Young's account, the two delusions differ in their experiential contents. According to Gerrans (2002), there is indeed a significant difference. The experience of Cotard patients is not restricted to faces but involves more generalised feelings of strangeness and unfamiliarity, culminating in an experience of disembodiment that is rationalized as the belief that one is dead, damned, or non-existent. Although Gerrans offers a one-stage explanation of the Cotard delusion, the view that the two delusions involve different perceptual contents is of course also compatible with two-stage explanations of one, the other or both.

All of these are 'bottom-up' approaches, which maintain that delusional beliefs originate in altered experience. In contrast to such approaches, Campbell (2001) offers a 'top-down' explanation of the Capgras delusion, according to which the altered experience is caused by the delusional belief, rather than being a cause of the belief. He appeals to the Wittgensteinian concept of a 'framework' or 'hinge' proposition, meaning a proposition that forms part of the background to explicit thought, as opposed to a proposition that is explicitly entertained and judged to be true or false.

A delusion, Campbell suggests, could be an anomalous framework proposition that leads to changed feelings and thus has an effect upon experience.

The contrast between these two alternatives is, I think, a misleading one. If we conceive of experience as the presentation of perceptual contents to a voyeuristic subject, then it might indeed seem that absence of the feelings that are usually associated with those contents is very different from a change in framework propositions. However, as recognised by phenomenologists such as Husserl, Heidegger and Merleau-Ponty, experience involves a great deal more than that. For example, healthy experience incorporates a background sense of being situated in a practically significant world. Alterations in feeling can have a profound effect upon this presupposed sense of belonging and consequently upon how objects of experience, including oneself and other people, appear. It is arguable that the experiential background does not differ so much from what Wittgenstein (1975) describes in terms of hinge propositions. For Wittgenstein, these are not 'propositions' in the usual sense of the term. Rather, they constitute a kind of practical orientation that shapes all experience and thought. Hence the distinction between top-down and bottom-up accounts relies upon an overly restrictive conception of experience, which drives the tendency to interpret changes in the experiential background as post-experiential 'cognitive' changes that may or may not feed back into experience in a 'top-down' fashion. In the remainder of this chapter, I will argue that the Capgras and Cotard delusions can be understood in terms of 'existential' changes in the background structure of experience.

Affect and Experience

It is widely acknowledged that diminished affect is largely responsible for changed experience in the Capgras and Cotard delusions. However, 'affect' can be understood in different ways. Some authors construe it phenomenologically, as something that is experienced. For example, in discussing the Capgras delusion, Ellis and Lewis (2001, p. 155) state that recognition of a familiar person can include an "automatic, concurrent 'glow'", thus indicating that the relevant affect is something we are aware of. In contrast, Davies et al. (2001, p. 140) treat 'affect' as a non-conscious, neurobiological process that sometimes has an effect upon experience. According to this latter view, Capgras patients do not register a change in conscious feeling but they do recognise that "something is different".

Capgras and Cotard sufferers complain of *feelings* of unfamiliarity, estrangement and unreality. Hence their reports strongly suggest that changes in conscious feeling have some role to play. And, as my emphasis here is upon phenomenology, I will focus on 'affect' understood as 'conscious feeling', rather on the role of unconscious affective processing.

There is the added complication of specifying just what impact changed feelings have upon experience. Consider the Capgras delusion. An absence of affect might involve seeing someone who looks just like the spouse and, at the same time, failing to have the feeling that is usually associated with such an experience. Alternatively,

the absence of affect could manifest itself as a change in how the spouse herself appears, rather than in an experiential change that is merely *associated* with her appearance. Perhaps she looks somehow strange or cold, no longer bathed in emotional warmth, oddly unfamiliar. A stronger claim is that the content 'this person is not my spouse' is intrinsic to the experience. It is hard to see how either an absence of those feelings ordinarily associated with the spouse or the spouse looking a bit strange could fuel the delusion. But the stronger claim has its problems too – it is not clear how diminished affect could constitute a changed experiential content. Analogous problems apply in the case of the Cotard delusion.

One objection to the view that feelings have a substantial impact upon experience is that they are merely perceptions of internal bodily states. In the Capgras delusion, at least, it is objects in the external world – principally other people – that look different. But this commonplace division of experience into perception of internal states and perception of the external world is something that can be challenged with the aid of phenomenology. Merleau-Ponty, for example, emphasises that the body does not usually feature as an *object* of experience. Rather, it is that through which we experience other things. All experience is structured by a background of bodily capacities and dispositions. It is the body that opens up a world for us; its dispositions constitute an orientation, a sense of being in the world, through which objects of experience are encountered:

> To have a body is to possess a universal setting, a schema of all types of perceptual unfolding and of all those inter-sensory correspondences which lie beyond the segment of the world which we are actually perceiving. (Merleau-Ponty 1962, p. 326)

According to both Husserl (1989) and Merleau-Ponty (1962), the world that we experience is not just a place of neutral actualities that are presented to our indifferent gaze but a realm of possibilities that are variably enticing. The things we experience are surrounded by salient possibilities for perceptual access and for practical manipulation by ourselves and others. Different objects, as experienced, call forth a range of different activities; these possibilities are experienced as residing in the things that we encounter. Husserl and Merleau-Ponty claim that the possibilities that are sewn into experience are constituted by bodily dispositions. The disposition to act in a particular way need not itself be the primary object of experience; it can be a way in which something else is experienced.[2] It is a short step from this view to the appreciation that something can be both a bodily feeling and a way of experiencing something outside of the body. Bodily feelings are surely inextricable from at least some of the bodily dispositions that structure world-experience.[3]

---

[2]This view is complemented by numerous recent findings in cognitive science, which similarly indicate that the body is not just an object of perception but a background framework that structures all perception. Gallagher (2005) brings together findings from phenomenology and science to argue at length for the view that the body has an enabling role to play in all perception: "In its *prenoetic* roles the body functions to make perception possible and to constrain intentional consciousness in various ways." (pp. 138–139)

[3]See my 'The Phenomenology and Neurobiology of Moods and Emotions' (this volume) and also Ratcliffe (2008a, Chapter 4) for more detailed discussions of feelings, possibilities and world-experience.

Once it is acknowledged that feelings need not be perceptions of a body-object, we can begin to see how they might play a role in delusional experience. However, what also needs to be recognised is that certain changes in feeling do not manifest themselves as ways in which particular objects or situations appear but as alterations in the overall structure of experience. This, I will suggest, applies to most or maybe even all delusions. And it certainly applies to the Capgras and Cotard delusions. In both cases, altered bodily feeling is also an altered way of being in the world, a change in felt connected to things, in the sense of significance, practical familiarity and reality.

Delusional Atmosphere

One of the reasons why the Capgras and Cotard delusions have received so much attention from neuropsychologists and neuropsychiatrists is that they seem to be fairly neat and tidy. They are monothematic, meaning that they have a specific and stable content, and they are circumscribed, meaning that they are seldom elaborated in intricate ways. So they look like better candidates for explanation in terms of particular patterns of cognitive breakdown than many of the other delusions that occur in illnesses such as schizophrenia, which are polythematic, elaborate and changeable. But the problem is that neither delusion is quite so tidy. The Capgras delusion is not usually the result of specific neurological damage and is more often found in psychiatric illnesses such as paranoid schizophrenia.[4] It is not always restricted to perception of familiar faces either. Some patients complain that possessions, pets or places have been replaced. And there is a lowering of affect in response to all faces, even though the absolute decrease is greater for familiar faces, which previously elicited the most pronounced response (Ellis and Lewis 2001, pp. 149–150). Furthermore, the delusion does not always involve visual perception. It has been reported in blind people, who maintain that voices have changed, and some sighted people also have it in relation to voices (Young and de Pauw 2002, p. 58). In addition to all this, patients almost always undergo a more general change in the shape of experience, which takes on an all-enveloping feeling of strangeness, unfamiliarity and unreality. It is not just people's faces that look odd; everything appears not quite right, distant and artificial:

> Although the impostor accusations may be directed at only one or two people, Capgras patients commonly report more pervasive feelings of strangeness, loss of affective response, and feelings that everything is somehow unreal or unfamiliar. (Stone and Young 1997, p. 337)

---

[4]Malloy et al. (1992) distinguish between primary and secondary Capgras delusions, where the former is associated with a psychiatric history, often schizophrenia, whereas the latter arises following neurological damage and, unlike primary Capgras, is not associated with paranoia.

> Patients sometimes report that everything looks strange, for example, things may look painted or not natural and faces may look like masks or wax models or seem to have been changed by plastic surgery. (Ellis and Young 1990, p. 241)

The Cotard delusion likewise arises in a range of psychiatric illnesses, including severe depression and, less often, schizophrenia. It too involves changes in the overall structure of experience. Patients' reports suggest a radically altered way of being in the world, where absent feeling is bound up with changes in experience of themselves, the world and the relationship between the two. They feel as though they are no longer animated, their bodies seeming mechanical and alien. In conjunction with this, the world no longer solicits activity. Everything is stripped of practical familiarity and significance. A sense of being part of the world, of inhabiting a place where things can show up as 'real', as 'present', as 'there', is gone. When patients complain of unreality, they are not complaining that specific objects in the world, which may include their own bodies, do not look real. Rather, the possibility of anything appearing as 'real' is absent from experience. Patients no longer feel that they are *there*, part of the world, connected to things. What motivates claims such as 'I am dead' is not an anomalous perceptual content but a changed way of finding oneself in the world, involving the erosion of a practical meaningfulness that we ordinarily take for granted as a backdrop to our experiences and thoughts. As Sass (2004, p. 74) remarks, the Cotard delusion incorporates a "transformation in what might be called the very framework of experience – in the normal, affectively grounded sense of existing as a living subjectivity". Gerrans (1999, pp. 603–604) likewise suggests that the Cotard patient has "effaced herself from the universe: nothing which occurs is of any significance to her and, hence, she describes the world without implicating herself in that description".

So, in both delusions, there is a change in the overall tone of experience, which is not restricted to only certain perceptual contents. I will now suggest that these delusions are not 'propositional attitudes' inferred from anomalous experiences. Rather, they are expressions of 'existential changes' in the structure of experience. These changes can be expressed in different ways, as involving aliens or robots, or being dead, damned or disembodied, for example. Hence expression also involves some degree of interpretation.[5]

Jaspers (1962) identifies a class of "primary delusions", meaning bizarre delusions that cannot be understood as arising from a history of meaningful events. Throughout his discussion, he places the emphasis upon existential changes, stating that a delusion is "a transformation in our total awareness of reality", rather than an experiential content that motivates a belief (p. 95). He refers to this transformation as a delusional *atmosphere* or *mood* and describes it as follows:

> [P]erception is unaltered in itself but there is some change which envelops everything with a subtle, pervasive and strangely uncertain light. A living room which formerly was felt as neutral or friendly now becomes dominated by some indefinable atmosphere. (1962, p. 98)

---

[5] Parnas and Sass (2001, p. 114) similarly claim that delusions are not so much "a matter of cognitive principles or intellectual rules so much as of what might be termed *general existential orientations*".

This 'atmosphere' is very difficult to describe, given that the actual features of perceived objects are no different. However, once it is acknowledged that experience also incorporates possibilities, we can see how two states of affairs can look the same and yet different. Although the actual characteristics of things might not change – colours, textures, shapes and relative positions – the possibilities that they offer up may still be different. For example, things might no longer appear enticing or tangible. And the world as a whole could be stripped of the practical significance that experience more usually takes for granted.

The Capgras and Cotard delusions are, I suggest, largely explicable in terms of changed possibilities that are integral to experience. I do not wish to maintain that the 'delusional atmosphere' has the same character in both cases. Indeed, I think it is important to recognise that the possibility space, the overall shape of experience, can be changed in all sorts of ways. For example, a sense of possibilities involving other people might be diminished or absent from experience. This is different from a sense of everything as intangible or distant, no longer offering up possibilities for practical manipulation. Intangibility is different from a general loss of practical significance, given that things can appear unobtainable whilst remaining significant. And a loss of significance differs from the world as a whole taking on the form of an all-encompassing threat before which one is passive, vulnerable and helpless. All of these are existential changes, alterations in the sense of reality and belonging that might constitute or contribute to a 'delusional atmosphere'.[6]

The Capgras delusion involves an absence of distinctively *personal* possibilities from experience, and the Cotard delusion a draining away of practical familiarity and significance, culminating in a loss of the feeling of belonging to a world. I remain neutral as to whether the two delusional experiences are qualitatively different or whether the latter is simply a more intense variant of the former. Given the conception of experience that I am advocating, both depressed mood and feelings of persecution are to be regarded as ingredients of experience, as opposed to being cognitive dispositions that are distinct from experience. They contribute to the overall shape of experience, which takes on the form of all-enveloping threat in one case and loss of all practical significance in the other. However, the difference between the two could be partly or wholly down to differing intensity. In the Cotard delusion, the diminishment of practical belonging might be so severe that the possibility of being threatened drops out. The patient becomes estranged from the world to such an extent that all self-concern vanishes and the feeling of threat is replaced by indifference.

How can the Capgras delusion be an expression of existential change, as opposed to a propositional attitude? Campbell (2001, p. 90) takes its content to be: "That [currently perceived] woman is not that remembered woman" and claims that this content cannot be constituted by feeling. However, he is wrong on both counts.

---

[6] See Ratcliffe (2008a, Chapters 5 to 7) for a discussion of some of the different ways in which the possibility space can be altered. See Ratcliffe (2008b) for a detailed account of existential changes in the Capgras delusion.

The actual content is not 'this perceived p is not that remembered q' but rather 'remembered person q has been replaced by p' or 'p is an *impostor*'. And this impostor content can, I propose, be understood in terms of changed feeling. As noted earlier, people with delusions often report a more general change in their experience – everything looks somehow fake or artificial. And Capgras sufferers frequently maintain that *people* look strange, impersonal or mechanical, like models.[7] It is as though a sense of others as *people* has been removed from experience. Phenomenologists, including Husserl (1989), Schutz (1967), Scheler (1954) and Sartre (1989), stress that we do not experience other people as complicated inanimate objects that are hypothesised to have internal mental lives somewhat like our own. We are immediately aware of others as loci of experience and agency. We encounter each other through what Husserl calls a "personalistic attitude", a practical, bodily orientation through which they are disclosed to us as animate agents. Our affective responses to other people play a significant role in enabling our experiences of them *as people*. Other people offer up distinctive kinds of possibilities for acting and being acted upon, including possibilities for expression, gesture, communication, forms of interaction and reciprocation. Our sense of these possibilities is inextricable from our felt bodily dispositions. Our bodies resonate to others in ways that are quite different from the ways in which they respond to other things.[8]

If the relevant kinds of possibility were absent from experience, if the feelings involved were diminished or gone, people would look very different, curiously impersonal. An impersonal entity that looked just like a person would strike one as a *fake* or *replica* person. And this is exactly what the Capgras experience seems to consist of. Here's how Renee describes this kind of experience in *Autobiography of a Schizophrenic Girl* (Sechehaye 1970):

> I knew her name and everything about her, yet she appeared strange, unreal, like a statue. I saw here eyes, her nose, her lips moving, heard her voice and understood what she said perfectly, yet I was in the presence of a stranger. (p. 36) I look at her, study her, praying to feel the life in her through the enveloping unreality. But she seems more a statue than ever, a manikin moved by mechanism, talking like an automaton. It is horrible, inhuman, grotesque. (p. 38)

The friend looks odd because a sense of people as *people* is gone from Renee's experience; the possibilities that are constitutive of a personal encounter no longer show up. Now this might not be so troubling if the perceived person is a shop assistant or somebody one passes on the street. One can buy groceries from others or dodge past them without worrying too much about whether or not they are

---

[7] See also Fuchs (2005) for a phenomenological discussion of delusional mood, which emphasises the experience of artificiality and estrangement.

[8] Accounts of interpersonal understanding in mainstream Anglophone philosophy are curiously oblivious to the fact that we encounter others as *people* rather than as *complex mechanisms*. See Ratcliffe (2007) for a critique of such accounts and also for a lengthy discussion of the phenomenology and neurobiology of *personal* experience.

people. But the situation is very different in the case of a spouse. A spouse who appeared slightly or perhaps even wholly impersonal, who no longer offered salient possibilities for affective relatedness, would look like a fake spouse, a replica spouse, a strange 'impostor' who appeared somehow more like an impersonal 'robot' or 'alien' than an ordinary person. Thus a sense of the perceived person being an impostor is part of the experience. It is not an isolated experiential content but an aspect of a more general experiential change. The change in question is not restricted to people either – there is a more general loss of practical significance and connectedness, with familiar objects and places seeming somehow distant, detached and devoid of purpose.

There is still the question of why the patient takes her experience to be veridical and adopts the belief that her spouse really is an impostor. But it would be a mistake to regard any anomalous reasoning as wholly separate from altered experience. The patient's sense of belonging to the world is radically changed and this will surely affect her reasoning too. Reasoning does not take place in some strange place outside of the world we find ourselves in. It too is embedded in our existential orientation. A richer conception of experience, drawn from phenomenology, challenges the dissociation between experience and reasoning that is assumed by two-stage models. And erosion of the sense that others are people will of course affect reasoning. We ordinarily acquire, regulate and change our beliefs in response to the testimony of others. But one would not be so inclined to seek or trust the testimony of others if one no longer experienced them as people. Hence much of the evidence required to challenge the belief is rendered inaccessible to the patient, who therefore accepts her experience as veridical.

Much the same approach can be applied to the Cotard delusion. It is a change in the overall structure of experience and claims such as 'I am dead' or 'I have ceased to exist' express this, rather than being propositional attitudes directed at an anomalous object of experience (oneself or one's body) that resides in an otherwise unadulterated experiential world. It might be objected that no experiential change warrants the claim that one is quite literally dead. Such utterances look to be blatantly self-contradictory – dead people can't tell you that they are dead. But the term 'dead' is used in different ways in different situations and need not be understood in a strict biological sense (Bayne and Pacherie 2005, p. 182). And it is by no means clear that the utterance 'I am dead' is self-contradictory – most people seem to have very little problem conceiving of ghosts, zombies and other species of 'living dead'. Furthermore, as Sass (2004) points out, talk of death is often associated with kinds of negative feeling. For example, people sometimes talk of feeling 'dead inside'. So using the term 'dead' to describe one's existential predicament is not necessarily problematic. However, a problem remains – there is a big difference between claiming that it is *as if* one is dead, disembodied or nonexistent and claiming that one is *actually* dead, disembodied or nonexistent. The 'as if' claim might well be used to communicate the kinds of existential change that I have referred to but Cotard patients go further than this, adopting the belief that they really are that way. Some account of the transition from 'as if' to 'belief' is therefore required.

I doubt that the shift from 'as if' to full-blown belief is so clear-cut. Jaspers (1962) describes the descent from depersonalisation and derealisation into a Cotard-like state, indicating in the process that there is continuity here rather than an abrupt transition from feeling to belief. Initially, "everything appears as though through a veil; as if I heard everything through a wall […] I touch myself to convince myself that I exist" (p. 62). As reality drains away still further, patients become alienated from social reality and from their own bodies, lost in a realm that offers only a painful sense of estrangement from everything:

> Terrified and restless, the patients begin to experience their feelings as the reality itself and are then inaccessible to reason. Now the world has escaped them. Nothing remains. They are alone in terrible isolation, suspended between infinities. They have to live for ever because time no longer exists. They themselves no longer exist; their body is dead. Only this fake-existence remains as their horrible fate. (p. 63)

Given the centrality of changes in the experience of reality, Jaspers claims that conceptions of delusions as mistaken and irrational 'beliefs' are "superficial and incorrect" (p. 93).

What patients' utterances express is an all-encompassing existential shift. This shift, as it becomes more pronounced, will inevitably have an impact upon reasoning. Reasoning is not insulated from how the patient now finds herself in the world – it does not occur outside of the world. As with the Capgras delusion, the patient becomes increasingly estranged from the social world in which reasoning more usually occurs and consequently from various interpersonal checking procedures.[9] Furthermore, it should not be assumed that she continues to 'believe' in the same *way* that she did before her existential orientation shifted. The loss of practical belonging, familiarity, significance and connectedness that Cotard patients describe is at the same time an erosion of the sense of reality. It is not simply that all experiential contents somehow appear as 'not real, as opposed to real'. Rather, the usual *sense* of reality that is presupposed by the *possibility* of anything appearing as 'real' is drained from experience. So, in asserting 'it is the case that p', where p is 'I am dead' or 'I have ceased to exist', the patient does not express the same kind or kinds of conviction that characterise instances of everyday belief. She cannot believe anything in that way anymore because the sense of reality that operates as a backdrop to everyday conviction is gone. She does not doubt that her predicament really is as it seems because that predicament is such as to estrange her from alternatives that would be presented to someone who still had a firm grounding in the social world. The patient has lost the sense of a public reality with which her experience of things might be contrasted and so the possibility of it only being *as if* she were dead is unavailable to her.

---

[9] I am grateful to Owen Earnshaw for pointing out to me the extent to which the reasoning processes that lead to delusional beliefs are interpersonal, as opposed to being processes that reside exclusively inside the heads of delusional people.

Nothingness

In addition to the absence of certain kinds of possibility, delusional experiences can also incorporate positive *feelings of absence*. For example, Capgras patients do not just *lose* the feeling of familiarity; "the normal feeling of familiarity has been replaced by a disturbing feeling of unfamiliarity and estrangement" (Bayne and Pacherie 2004, p. 4). Patients appreciate that there is something not quite right about their experience; something is missing, unpleasantly lacking. The Cotard delusion similarly involves an all-encompassing experience *of* absence.

This aspect of experience is not explained by accounts that emphasise propositional attitudes. When you negate proposition p, you do not have an experience of nothingness or of the absence of p. Propositional negation has no phenomenology. In contrast to such approaches, Sartre and Heidegger both claim that nothingness does not originate in propositional negation. In fact, a sense of what it is for something not to be is presupposed by the possibility of propositional negation. As Sartre (1989, p. 7) puts it:

> Is negation as the structure of the judicative proposition at the origin of nothingness? Or on the contrary is nothingness as the structure of the real, the origin and foundation of negation?

He opts for the latter (as does Heidegger). Experienced negation, for Sartre, is facilitated by the incorporation of possibilities into experience: "The world does not disclose its non-beings to one who has not first posited them as possibilities" (p. 7). A situation is configured in terms of salient possibilities, some of which take the form of explicit or implicit expectations. Where an experience is shaped by expectations and what is expected fails to occur, its absence is a real part of the experience.

Sartre offers the famous example of going to the café to meet Pierre. When Pierre fails to show up, his absence from the situation, his non-being, permeates experience of the café, which becomes a background to a non-existent foreground, analogous to an elaborate picture frame with nothing in it. The manner in which the café appears is structured by something that is not there; the absence haunts the experience as a whole. World-experience, Sartre says, is riddled with these "little pools of non-being" (p. 19). Of course, this applies to only some absences. Pierre is absent from the café in a way that Wellington is not. The propositions "Pierre is not in the café" and "Wellington is in the café" are equally true of the scenario in question but only the latter involves an experience of absence. Hence an emphasis on which propositions are accepted fails to convey the relevant experience.

I think that something not unlike this occurs in the Capgras delusion. Unlike Pierre, the spouse does show up. But, although he is present, there is a feeling of unfulfilled expectation. When he smiles, the smile does not offer up the usual possibilities for communication, reciprocation and so on, possibilities that are constitutive of experiencing him as a person. So the experience incorporates a sense of *personal* expectations being unfulfilled, of the personal being conspicuously and consistently absent. There is the question of what form the relevant expectations

take, of whether the Capgras patient continues to feel for the remembered spouse and expects to experience the perceived spouse in the same way or whether she merely remembers feeling for the remembered spouse and expects to experience the perceived spouse in that way. Both are compatible with unfulfilled expectation, with what is now being perceived not *looking how it is supposed to look*. However, the latter is more likely – Capgras patients have diminished feeling and are also frequently reported to display a lack of concern for the missing spouse, suggesting that the person as perceived and as remembered both fail to elicit an affective response.

The Cotard delusion is more akin to something that Heidegger describes. Heidegger offers an account of mood [*Stimmung*] as a background sense of belonging to the world, which we ordinarily take for granted. Certain unusual moods involve, he suggests, a draining of significance from the world and an erosion of belonging. The world no longer appears as a realm of practical possibilities but as a strange place from which one is disconnected. It is encountered through an all-embracing feeling of absence. The culmination of this is a mood that he calls *Angst*, in which "the world has the character of completely lacking significance" (Heidegger 1962, p. 231). All that remains is a feeling of having no practical hold on things, no connectedness to anything. Experience as a whole is characterised only by a sense of utter loss, by the world's failing to offer up anything at all. It "can offer nothing more" (p. 232) and "everyday familiarity collapses" (p. 233). As Heidegger recognises, the kind of felt familiarity with which experience is ordinarily imbued is not just a warming buzz associated with certain objects of perception; it is an orientation that makes it possible for things to show up as significant, as real even. When it falls away, all that remains is what Heidegger refers to as the "nothing", an experienced loss of the familiar, everyday world, "the complete negation of the totality of beings" (Heidegger 1978, p. 98).

I suggested earlier in this chapter that a delusional atmosphere might well take a variety of subtly different forms and I certainly do not want to claim that the Cotard delusion has exactly the same phenomenology as Heideggerian Angst. Indeed, I doubt that Heidegger's references to 'Angst' succeed in identifying a single, distinctive kind of existential predicament (see Ratcliffe 2008a, Chapter 2). Nevertheless, the work of phenomenologists such as Heidegger is instructive in its acknowledgement that there is more to nothingness than propositional negation. The Cotard experience incorporates an all-encompassing apprehension of nothingness or non-existence, and the acknowledgement that absence and negation can be part of an experience is required in order to adequately interpret it.

Conclusion

I have argued that the Capgras and Cotard delusions can be understood as expressions of existential orientations, rather than as propositional attitudes that are caused, at least in part, by specific experiential contents. In both cases, the experiential changes

can be expressed in different ways, as involving aliens, robots, death, disembodiment or damnation, for example. However, not all delusions can be understood in the same way. Many of the delusions that arise in schizophrenia take the form of elaborate and changeable narratives, involving a range of quite specific contents.[10]

It is implausible to suggest that delusions of being chased by secret agents and the like are *expressions* of how one finds oneself in the world. Nevertheless, it is likely that such delusions are at least *symptomatic* of existential orientations. They are not simply propositional attitudes formed against the backdrop of an otherwise ordinary sense of reality but are cultivated by global changes in the structure of experience. As with the Capgras and Cotard delusions, the relevant changes involve a diminution, absence or alteration of feeling. This is something that Louis Sass has stressed in various works. He argues that it is a mistake to distinguish 'negative' symptoms in schizophrenia such as diminished affect from 'positive' symptoms such as delusions. Diminished affect is indissociable from changes in the structure of experience and these changes themselves comprise an existential background that is conducive to the formation of delusions. The relationship between these factors is not to be construed mechanistically. Rather, it is one of "mutual phenomenological implication" (Sass 2003, p. 157).

According to Sass (e.g. Sass 1994), 'hyperreflexivity' plays a central role in schizophrenia. This, he explains, is a disposition to reflect excessively upon what is more usually taken for granted as a *medium* of experience, thought and activity. That through which we ordinarily experience and belong to the world becomes object-like. The resultant predicament involves an erosion of practical connectedness to the world, a fragmentation and loss of experienced functionality, an alienation from one's own body and a retreat to a somewhat solipsistic experiential realm, far removed from the social reality that we ordinarily presuppose as a backdrop to our lives. As Parnas and Sass (2001, p. 105) put it, "the patient does not feel being fully existing or alive, fully awake or conscious, or fully present and affected". Her perception is "not lived but is more like a mechanical, purely receptive sensory process, unaccompanied by its affective feeling-tone". It is against the backdrop of this altered experiential realm that delusional narratives are cultured.

The question remains as to how many variants of delusional atmosphere there are. As indicated earlier, I suspect that there are many and that further phenomenological enquiry into the *kinds* of modification that our sense of being in the world can undergo is needed in order to determine whether and how the experiences implicated in different delusions differ from each other.[11] But, even without the complete picture, it is at least clear that phenomenological enquiry has an important role to play in understanding delusional experience. Accounts that emphasise specific experiential contents and construe them as a stage in the process of propositional attitude

---

[10] See, for example, the delusional universe related by Daniel Paul Schreber in his *Memoirs of my Nervous Illness*.

[11] See Ratcliffe (2008a) for a description of several such predicaments and for the view that existential orientations are at the same time bodily feelings.

formation neglect the fact that all experiences also incorporate ways of finding ourselves in the world, backgrounds within which we think and act. Changes in the experiential background are largely responsible for many and maybe all of those phenomena that are referred to as 'delusions'. Delusions are not propositional attitudes that involve taking an unreal thing to be real. Some are ways of being in the world and others are symptomatic of ways of being in the world.

References

American Psychiatric Association (2004) Diagnostic and statistical manual of mental disorders, 4th edition, text revision. American Psychiatric Association, Washington, DC

Bayne T, Pacherie E (2004) Bottom-up or top-down? Campbell's rationalist account of monothematic delusions. Philos Psychiatry Psychol 11:1–11

Bayne T, Pacherie E (2005) In defence of the doxastic conception of delusions. Mind Lang 20:163–188

Bortolotti L (2005) Delusions and the background of rationality. Mind Lang 20:189–208

Broome MR (2004) The rationality of psychosis and understanding the deluded. Philos Psychiatry Psychol 11:35–41

Campbell J (2001) Rationality, meaning, and the analysis of delusion. Philos Psychiatry Psychol 8:89–100

David AS (1999) On the impossibility of defining delusions. Philos Psychiatry Psychol 6:17–20

Davies M, Coltheart M, Langdon R, Breen N (2001) Monothematic delusions: towards a two-factor account. Philos Psychiatry Psychol 8:133–158

Ellis HD, Young AW (1990) Accounting for delusional misidentifications. Br J Psychiatry 157:239–248

Ellis HD, Lewis MB (2001) Capgras delusion: a window on face recognition. Trends Cogn Sci 5:149–156

Fuchs (2005) Delusional mood and delusional perception – a phenomenological analysis. Psychopathology 38:133–139

Fulford KWM (1994) Value, illness and failure of action. In: Graham G, Stephens GL (eds) Philosophical psychopathology. MIT Press, Cambridge, MA, pp 205–233

Gallagher S (2005) How the body shapes the mind. Oxford University Press, Oxford

Gerrans P (1999) Delusional misidentification and subpersonal disintegration. The Monist 82:590–608

Gerrans P (2002) A one-stage explanation of the cotard delusion. Philos Psychiatry Psychol 9:47–53

Heidegger M (1962) Being and time (trans: Macquarrie J, Robinson E). Blackwell, Oxford

Heidegger M (1978) What is metaphysics? In: Heidegger M (ed) Basic writings (trans and ed: Krell DF). Routledge, London, pp 93–110

Husserl E (1989) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: second book (trans: Rojcewicz R, Schuwer A). Kluwer, Dordrecht

Jaspers K (1962) General psychopathology. Manchester University Press, Manchester

Maher B (1999) Anomalous experience in everyday life: its significance for psychopathology. Monist 82:547–570

Malloy P, Cimino C, Westlake R (1992) Differential diagnosis of primary and secondary capgras delusions. Neuropsychiatry Neuropsychol Behav Neurol 5:83–96

Merleau-Ponty M (1962) Phenomenology of perception (trans: Smith C). Routledge, London

Parnas J, Sass LA (2001) Self, solipsism and schizophrenic delusions. Philos Psychiatry Psychol 8:101–120

Ratcliffe M (2007) Rethinking commonsense psychology: a critique of folk psychology, theory of mind and simulation. Palgrave Macmillan, Basingstoke

Ratcliffe M (2008a) Feelings of being: phenomenology, psychiatry and the sense of reality. Oxford University Press, Oxford

Ratcliffe M (2008b) The phenomenological role of affect in the capgras delusion. Continent Philos Rev 41:195–216

Sartre JP (1989) Being and nothingness (trans: Barnes HE). Routledge, London

Sass LA (1994) The paradoxes of delusion: Wittgenstein, Schreber and the schizophrenic mind. Cornell University Press, Ithaca

Sass LA (2003) 'Negative symptoms', schizophrenia, and the self. Int J Psychol Psychologic Therapy 3:153–180

Sass LA (2004) Some reflections on the (analytic) philosophical approach to delusion. Philos Psychiatry Psychol 11:71–80

Scheler M (1954) The nature of sympathy (trans: Heath P). Routledge, London

Schreber DP (2000) Memoirs of my nervous illness. New York Review of Books, New York

Schutz A (1967) The phenomenology of the social world (trans: Walsh G, Lehnert F). Northwestern University Press, Evanston

Sechehaye M (1970) Autobiography of a schizophrenic girl. Signet, New York

Stone T, Young AW (1997) Delusions and brain injury: the philosophy and psychology of belief. Mind Lang 12:327–364

Wittgenstein L (1975) On certainty (trans: Paul D, Anscombe GEM). Blackwell, Oxford

Young AW, Leafhead KM (1996) Betwixt life and death: case studies of the cotard delusion. In: Halligan P, Marshall J (eds) Method in madness: case studies in cognitive neuropsychiatry. Psychology Press, Hove, pp 147–171

Young AW, de Pauw KW (2002) One stage is not enough. Philos Psychiatry Psychol 9:55059

Autoscopy: Disrupted Self in Neuropsychiatric Disorders and Anomalous Conscious States

Aaron L. Mishara

Background: Self and Neuropsychiatric Disorders

Autoscopy is a loosely related complex of experiences in which one sees (or experiences) a "double" as external to one's current vantage point. The phenomenology of autoscopy provides an alternative to current "models" of self in cognitive science. After years of silence on matters such as consciousness and self, cognitive science and neuroscience have now swung in the opposite direction, and claim to be able to experimentally study these topics, often in an oversimplified manner. These approaches uncritically confuse representational content about self or self-awareness in self-referential processing, i.e., having a self (a self-enclosed entity), with being a self, prospectively *open* to its own (yet-to-be-known) future.[1] Ignoring this difference has led to an industry of philosophical essays and neuroimaging studies that claim to access the first-person perspective when only able to access higher order self-referential judgments (for critical reviews, see Fuchs 2006; Legrand et al. 2003; Mishara 2007b). A similar confusion prevails in current approaches to classify types of autoscopy in the search for its underlying cognitive-neural mechanisms.[2] Due to its

A.L. Mishara
Department of Psychiatry, Clinical Neuroscience Research Unit, Yale University School of Medicine, CMHC 339-A, 34 Park St., New Haven, CT, 06519, USA
Current Address: Department of Psychiatry, Brain Mapping Unit and Behavioural and Clinical Neurosciences Institute, University of Cambridge, School of Clinical Medicine, University of Cambridge Department of Psychiatry, Brain Mapping Unit, Addenbrooke's Hospital Cambridge, CB2 2QQ, UK

[1] By overlooking this important conceptual distinction, self-reference is confused with "being a self" (e.g., Gusnard 2005). By requiring subjects to make explicit, reflective judgments about their emotional experience, for example, they are required (ironically) to take an external, third-person perspective to their experience (See Mishara 2005, 2007b).

[2] Although the term "neurocognitive" is in vogue, I deliberately do not use it. It circumvents the mind-body (or mind-brain) problem or simply, treats it as already solved. Although associated with one another (as in lesion or neuroimaging studies), cognitive and neural are different levels of explanation (see Mishara 2007b). According to the "revolving door principle" (von Weizsäcker (1948), whose work is described in this contribution and elsewhere (Mishara 1994)), our access to cognitive (mind) and neural (brain) remains dual: cognitive and neural presuppose but also exclude one another (as two sides of a coin) in terms of a "hidden unity." We will examine how this "hidden unity" is disrupted in autoscopy.

descriptive method and resulting theoretical framework, phenomenology is in a unique position to contribute to the study of human self and its disruption in neuropsychatric disorders.

Classification of Autoscopy

The Four Types

Autoscopy (from the ancient Greek, 'seeing oneself') is the experience of seeing one's face, upper or whole body in external space from within one's physical body (autoscopy/heautoscopy) or from an external point of view (out of body experiences, OBEs). The "double" may also be felt (but not seen) in peripersonal space in the feeling of a presence (FOP). That is, autoscopy comprises a tetrad of loosely related experiences (seeing oneself, OBE and FOP). These experiences are sometimes co-occurrent in some disorders or anomalous conscious states but are thought to involve different underlying cognitive-neural mechanisms.

Descriptions of autoscopy date back to Aristotle. Medical accounts of autoscopy, however, began in the mid-nineteenth century (Wigan 1844). *Fin de siècle* treatments are the first to approach the topic systematically by introducing the descriptive terms, "hallucinations autoscopiques ou spéculaires" (Féré 1891; Sollier 1903). Literary authors often portray autoscopic experiences (e.g., G. D'Annunzio, F. Dostoevsky, J.W. v. Goethe, E.T.A. Hoffmann, G. de Maupassant, A. de Musset, E. A. Poe, J.P. Richter, P. B. Shelley, R.L. Stevenson), many of whom experienced autoscopy themselves (see Todd and Dewhurst 1955, for review).

The etiologies of autoscopic hallucinations are diverse and unrelated. They occur in epilepsy, brain tumors, labyrinthine vertigo, schizophrenia and depression. They also occur during drug intoxication, trauma-related dissociative experiences, the hypnagogic/hypnopompic hallucinations associated with sleep paralysis, and in individuals with high fantasy proneness. If the patient believes that a true double exists, even though invisible, or the hallucinations are experienced as a "true double," there is the added feature of the *delusion* of having a double (Christodoulou 1978). The autoscopic experience may last from seconds to hours or, in some cases, be present for years at a time (Conrad 1953; Engerth and Hoff 1929). From its first emergence, there is something about the double that fascinates (i.e., captures the attention). Although the identification of the double may not be immediate, its recognition (the "aha") is sudden (even if only viewing the double's back, e.g., Blanke et al. (2004), Devinsky et al. (1989)). If it appears to an otherwise clear consciousness (especially the purely visual kind, i.e., Type I, see below), it may lead to a perplexed reflectiveness (Menninger-Lerchenthal 1961).

In the mid 1990s, Denning and Berrios (1994) complained that autoscopy-research as a field lacked conceptual clarity. Subsequently, clinical-researchers (e.g., Blanke and Mohr 2005; Brugger 2002; Brugger et al. 1997; Lopez et al. 2008) proposed taxonomies of autoscopy, usually overlapping. As is evident from these

efforts, the categories used for classification are highly interrelated with the proposed cognitive mechanisms and underlying neurobiology. The taxonomies, however, may be more clear-cut than the actual phenomenology (e.g., Case 3, below). In Table 1, I present a slightly modified classification which lends itself to the subsequent phenomenological analysis.[3]

Type I Autoscopy: The Other Is Like Me

Case Examples:
Case 1

Conrad (1953) describes a 41-year-old man who became blind due to an untreated tumor of the hypophysis (pituitary gland). Eight days after a shocking and unexpected onset of blindness, the patient awakens to see a face (1 m away) looking down on him. "He is startled, and thinking at first that someone is there, reaches out to grasp it but reaches into emptiness. … The face, however, continues to appear before him. … A few days later, he realizes that it is actually his own face … It imitates all his facial movements *exactly as in a mirror*." (p. 266, my translation and emphasis). The condition has remains unchanged for the past 7 years.
Case 2

Suffering from terminal syphilis, the French writer, de Maupassant asks his friend, Bourget: "How would you feel if you had to go through what I experience? Every other time when I return home I see my double. I open the door and see myself sitting in the armchair. I know it is a hallucination the moment I see it. But isn't it remarkable? If you hadn't a cool head wouldn't you be afraid?" (Todd and Dewhurst 1955, p. 48).

Table 1 The four types of autoscopy

---

(1) Type I: Visual hallucinatory autoscopy
  – I is mirrored by a me (body or self as object)
(2) Type II: Delusional (dream-like) autoscopy (usually called heautoscopy)
  – I becomes a me, i.e., the mirror image (ironically) of the other I who usurps the feeling of being a self
(3) Out of body experience (OBE)
  – The I separates from the physical body and views it from an elevated position: I (body as subject) and me (body as object) are experienced as separate
(4) Feeling of a shadowy presence
  – Another I is sensed but not seen

---

[3] The more recent taxonomies of autoscopy (Blanke and Mohr 2005; Brugger 2002; Brugger et al. 1997; Grüsser and Landis 1991; Lopez et al. 2008) draw from the earlier classificational systems (Conrad 1953; Hécaen and Ajuriaguerra 1952; Lhermitte 1939; Menninger-Lerchenthal 1935; Sollier 1903). The major difference between the classification I propose here and those preceding is that I label the second type of autoscopy, "delusional (dream-like) autoscopy," often labeled heautoscopy (Brugger et al. 1997; 2002; Blanke and Mohr 2005; Lopez et al. 2008). See Tadokoro et al. (2006) for a different classification system.

Type I autoscopy is predominantly a unimodal, visual hallucination of the double's face, upper or entire body. It is frequently associated with lesions of occipital cortex and accompanying hemianopnia. Confined to the area of blindness, it often appears in lateral space. While these unbidden apparitions are life-like in color and often distressing, such "pseudohallucinations" are rarely accompanied by a belief in the reality of the experience (see Table 2).[4]

When the double exhibits echopraxia (ongoing, simultaneous imitation of the patient's movements in real time) (as in Case 1), it acts with the same left-right reversal that one experiences in a mirror. That is, the double's movements are symmetrical to the subject's; for example, the double uses the symmetrical (left) rather than contralateral hand to imitate the subject's movement with the right hand. As Conrad's patient, some patients see the mirror image in every direction they look (e.g., Kamiya and Okamoto 1982, Case 3).

The phenomenologically-oriented psychiatrist, Conrad (see Uhlhaas and Mishara 2007 and Mishara in press a; in press b, for reviews), doubts that the mirror-face was the product of optical hallucinations or denial of blindness (Anton's syndrome). Comparing it to the phantom limb, he states that it reflects an effort to retain the integrity of the body after loss, in this case, after the loss of vision. The threatened integrity directly involves the conviction that one exists. Although arising subpersonally from the internal sensations of the body, this conviction (for Conrad 1933; 1953) is also at the root of the most personal sense of self: the internal perception of our own bodies as existing (i.e., "coenesthesia") (see Mishara 2005 and "Dissociating Mind and Subjectivity," below).[5]

Type II Autoscopy: I Am Like the Other

Case 3

After visiting the grave of her recently deceased husband, a 56-year-old, retired schoolteacher returns home. Upon opening the door, she senses that someone else is in the house occupied only by her. In the twilight-lit room, she sees that another woman is standing in front of her. As she lifts her right hand to turn on the electric

---

[4] Unlike hallucinations proper, the person experiencing "pseudohallucinations" retains insight that the experience is unreal (Sedman 1966).

[5] The term "coenesthesia" was common until the beginning of the twentieth century when it was replaced by terms such as "body schema" and "body image" (e.g., Head and Holmes 1911; Head 1920; for Sir Henry Head's influence on German Phenomenological psychiatrists, especially Klaus Conrad, see Mishara in press a). It derives from the Aristotelian doctrine that information from the exteroceptive senses (vision, audition, touch, smell) only reaches 'internal sense' by means of common sense (*sensus communis*, *koinon aesthesis*), therefore the German, *Gemeingefuehl* (see Fuchs 1995). The sense of bodily existence, i.e., the very feeling that one is alive, that one's body exists, lies at the core of the human experience of self. This feeling of being alive subserved by an interoceptive neural pathway (and related reward-emotion processing) gives rise to the "myness" of my experience and may be disrupted in schizophrenia and other neuropsychiatric disorders (see Mishara 2004).

Table 2 Key features of Type I and Type II autoscopy

Type I: Visual hallucinatory autoscopy
- Predominantly visual
- Without accompanying somaesthetic experience
- Echopraxia (when present): double mirrors
- Allocentric coordinates of body image (double mirrors with symmetrical body)
- (Often) associated with lesions of occipital cortex and accompanying hemianopnia
- Visual hallucination may be of the face, upper or entire body
- Often lateralized and confined to the area of blindness
- Unbidden apparitions life-like in color
- Rarely accompanied by a *belief* in the reality of the experience i.e., "pseudohallucinations

Type II: Delusional (dream-like) autoscopy
- Multimodal
- Altered bodily awareness
- Depersonalization: feeling of belonging to double(s)
- Lightness, vestibular illusions, detachment, hollowness or splitting of one's own body (something vital has left the body)
- Proprioceptive–kinaesthetic disturbance associated with a strong psychological affinity between physical and autoscopic body
- A "projection of the body's visceral and deep sensations in the space on the outside of the body" (Sollier 1903; see Blanke and Mohr 2005)
- Echopraxia (when present): double mimmicks, anticipates or acts independently
- Egocentric (motoric) coordinates of body schema (double mimicks with contralateral side of body)
- Despite possible variation in appearance between the person and the double, the person somehow "knows" all at once the image to be the "double" (right away or after an initial brushing)
- There may be sharing of thoughts words or action. The patient may hear the autoscopic body talk to them (Brugger et al. 1994) or that both bodies communicated by thought (Blanke et al. 2004). *The autoscopic body may be performing the actions they were supposed to do* (Devinsky et al. 1989) or fights with other people that could be of potential danger to the patient (Blanke et al. 2004), or the subject fights with the double to the point of suicide or death (Brugger et al. 1994).

Type I vs. Type II Autoscopy
- Type I predominantly occipital lesion site, the autoscopic body usually more brightly colored > Type II Dreamlike Delusional double, often reported as colorless, pale, transparent, misty, ghost-like, or jelly-like (Feinberg 2001)
- Type II experienced as more real > Type I
- Importantly, the seeming reality of the hallucination does not depend on the characteristics of the spectre but on how the subject constitutes the experience, i.e., the subject's state of mind as dream-like

light, the figure makes the same movements with her left hand so that their hands meet. She remarks that her own hand feels cold and bloodless from the contact (paraphrased from Lukianowicz 1958).

Case 4

Report of 21-year-old patient with left temporal lobe epileptogenic focus and right temporal epileptiform discharges (revealed by interictal scalp-EEG):

"… On the respective morning he got up with a dizzy feeling. Turning around, he saw himself lying in bed. He became angry with 'this guy who I knew was myself and who would not get up'… The lying body showed no reaction. Only then did the patient begin to become puzzled about his double existence and become more and more scared by the fact that he could no longer tell which of the two he really was. Several times his body awareness *switched* from the one standing upright to the one still lying in bed; when in the lying in bed mode he felt quite awake but completely *paralyzed* and scared by the figure of himself bending over and being him. His only intention was to become one person again and, looking out the window (from where he could still see his body lying in bed), he suddenly decided to jump out 'in order to stop the intolerable feeling of being divided in two.'" Surviving the fall, the patient reports this experience in retrospect (Brugger et al. 1994, my emphases).

Unlike the mirroring in Type I autoscopy, Type II is characterized by the feeling that one (ironically) becomes the mirror image of the double, who usurps the feeling of being the "real self." There may be a feeling of oneness with the hallucination as if self and double are "emotionally linked" (Lhermitte 1951) share a "feeling of belonging" (Damas Mora et al. 1980), or complete one another (e.g., Kafka 1979; Richter 1959).

Type II autoscopy is multimodal. The double may speak with, or touch (e.g., Case 3) the subject. One's own bodily feelings, usually indicating depersonalization (e.g., disturbances in the experience of gravity, vestibular sensations, kinesthetic-proprioceptive abnormalities, unusual lightness) contribute to the double's convincing status. One's own body may be experienced with detachment, as hollow or light. The degree to which the double is experienced as convincing is inversely proportional to the feeling of depersonalization in oneself. "The subject may wonder whether it is the body or rather the Doppelgänger which contains the real self" (Brugger 2002, p. 184).

Although recognized to be the double, the apparition may vary in appearance from the subject, including clothes (Todd and Dewhurst 1955), age (younger or older), or even gender.[6] The double may be motionless, expressionless, imitate the

---

[6] Whereas Type I patients see the autoscopic body in front-view, Type II patients sometimes report seeing the autoscopic body in side- or back-views. Ionasescu (1960) reports a hairdresser, who saw his own autoscopic body from the side while cutting hair (cited by Blanke and Mohr 2005). Brugger (2002) interprets non-frontal Type II autoscopy (i.e., *heautoscopy*) to be computationally less demanding than full frontal views for the visual and other neural systems supporting the hallucination: "This type of visualization may spare the cognitive system the trouble of mentally 'turning around' in order to maintain consistency in the sidedness of the original and duplicated bodies" (p. 187). However, Brugger and colleagues (2006) report a case of "polyopic heautoscopy" (in which more than one double is experienced at the same time) which suggests that such reductions in computational complexity are not always present in Type II autoscopy. Upon seeing the first double, the patient reports his efforts of trying to discover its identity: "When I walked around, I repeatedly looked towards the gentleman on my side and wondered if I could recognize his face. This was impossible since on looking towards the right side he also turned his head to the right." (p. 669). This suggests that rather than reducing the computational complexity, the *Doppelgänger* – by precisely mimicking the patient's own efforts to discover the double's identity – not only evades the patient's efforts but does so with an exquisite precision that reproduces the *egocentric computations* and timing of the subject's own head movements!

subject's own facial expressions and movements (heutoscopic echopraxis), or act independently.[7]

The Type II double is often reported as colorless, pale, transparent, cloudy, misty, ghost-or jelly-like (see Feinberg 2000). In contrast, the Type I autoscopic body is – consistent with the predominantly occipital lesion site – generally more vivid and brightly colored. However, the Type II double is experienced as more convincing than its Type I counterpart. *The experienced "reality" of the hallucination does not depend, then, on the phenomenological characteristics of the specter but on* how *the subject constitutes the experience* (see Table 1). Type II autoscopy is more dreamlike, dependent on the actual state of consciousness of the person. Therefore, I have characterized this type as delusional (dream-like) autoscopy. The view that Type II autoscopy may be delusional (or dreamlike)[8] is supported by the fact that this form of autoscopy can lead to death or suicide through fighting with or trying to free oneself from one's double (e.g., Case 4).

Notably, the patient in Case 4 reports that the double attacked the subject with the same right-handed preference with which the subject defended himself (Brugger 2002). This indicates that the double was not cast as a mirror image (i.e., in terms of an allocentric frame of reference as in Type I mirror-like autoscopy) but seemed to employ the same egocentric coordinates of agency as the subject. That is, the double mimics the subject's actions contralaterally (i.e., not symmetrically as looking in a mirror but as if another self, in rotated perspective, were performing the same movements). The double may execute activities that the subject was just about to enact. It may "anticipate the person's actions, thoughts or words" (Todd and Dewhurst 1955). *This suggests that the double has access to the subject's own intentions before or during their execution.* Unlike Type I autoscopy

---

[7] In E.T.A. Hoffmann's (1967) remarkable short story*, New Year Eve's Adventure (Die Abenteuer der Sylvester-Nacht)*, the narrator – following a painful romantic disappointment at a New Year's party – finds himself in a smoky, basement pub, with two new acquaintances, one who has lost his mirror image in a similar romantic disappointment and another man who has unhappily sold his own shadow. Unable to go home, the narrator is mistakenly given the same hotel room as the man who has lost his mirror image. In the morning, he wakens to find that his unwanted roommate has already departed but has left behind a manuscript in which the author (named Erasmus) describes *in 3rd person* how he lost his mirror image. Critical for our discussion of Type II autoscopy is Erasmus' depiction of how the mirror image becomes detached and acts independently of the subject's intended movements: "Erasmus saw his image step forward independent of his movements, glide into Giuletta's arms and disappear in a vapor" (Hoffmann 1967, p. 122). For the justification of using literary depictions as evidence for (abnormal) structures of self, see phenomenological method, step 2, below.

[8] The "classical" psychopathological view (e.g., Jaspers, Gruhle, K. Schneider) is that "delusions can only arise in the process of thinking and judging." In contrast, some phenomenological-psychiatrists (e.g., Binswanger, Conrad, Blankenburg) hold the view that delusions are related to more automatic processes, not unlike dreaming (Mishara 1997; Uhlhaas and Mishara 2007). Conrad (1958; 1960) believes that the delusional schizophrenia patient is "caught between waking and sleeping" see Mishara in press a; in press b.

(in which one's mirror or body image is implicated), one's agentic self (i.e., body schema) is impacted.[9]

Thus, Type II autoscopy is distinguished from Type I by (a) the extent to which the patient believes the image to be a double of the self; (b) echopraxia (when present) is experienced in egocentric as opposed to allocentric coordinates (thus reflecting body schema rather than body image, as will be further explained below); and (c), the experience is multimodal, accompanied by bodily sensations (i.e., proprioceptive-kinesthetic and vestibular sensations) which include weakness, feelings of hollowness and/or depersonalization of the original bodily self.

Out of Body Experience (OBE): The Splitting of "I" and Me

Case 5

A 29-year-old man, with history of seizures (progressing to loss of consciousness) since the age of 12, reports that during the seizures, he ascends "… to a corner of the room. From there he could look down on his body, while feeling very cold and lightheaded. The body below lay motionless on the floor or bed, while the 'mind above' could move around the house and see the other family members in the various rooms. He clearly saw the body below; it was not transparent and did cast a shadow. The body was dressed in the same clothes he was wearing, but curiously always had combed hair even when he knew his hair was uncombed before the onset of the episode. … He had these out-of-body episodes with almost every aura and felt that they were the most distressing component of the seizures" (Devinsky et al. 1989, p. 1081).

Grüsser and Landis (1991) define out of body experiences (OBEs) in terms of (1) "the splitting of 'I' and 'me'" (borrowed from a patient's description); (2) "the vestibular sensation of being elevated outside one's body" (p. 299). In OBEs, the point of view is usually elevated above one's visible body, which (when seen) is usually inert and lying down (Blanke and Mohr 2005). It may be experienced as

---

[9]When we view an image in a mirror parallel to one's own face and body, we are aware of a left-right but not a top-bottom reversal (Navon 2001). As a result, there is considerable discussion to what extent the "illusory" experience of self in the mirror image is due to top down cognitive processes or the bottom up processing of the optical geometry (see Navon 2001, and commentaries). Because the experience of the mirror occurs in conscious visual perception, it is experienced as a doubling of the visible body or body image. What is often overlooked, the mirror image is computed – like Type I autoscopy – in allocentric, object-centered coordinates. Lacan (1977) and others following (e.g., Merleau-Ponty) point to the following phases in the experience of mirror reflection: (1) a perceptual captivation in the mirror reflection; (2) a motoric phase in which the subject (initially the infant) takes glee in having the mirror image follow the movements which he or she initiates, (3) An anticipated, "illusory" totality of one's self integrates the perceptual and motoric moments or 'selves'.

(1) a (self-moving) *bodiless* point in space (e.g., a "ball of light," Green 1968); or (2) within an imaginary, weightless *replica of one's own body*, i.e., a "parasomatic," "astral" or "phantom body"[10] (see Table 2). One report testifies to how these two types of point of view may alternate: "… once out in a bodiless state, I caught myself beginning to assume a shape similar to my physical body because of familiarity with it" (Donahu 1974, p.61, cited by Irwin 1985, p. 115). As Type II autoscopy, the experience is frequently multimodal with auditory (e.g., buzzing, snapping or ringing noises (Jantz and Beringer 1944)), and vestibular-proprioceptive contributions especially during the experience of movement. OBEs are reported during the hynagogic/hypnopompic hallucinations of sleep paralysis, dreaming, extreme physical exertion (e.g., marathon running or high altitude climbing (Brugger et al. 1999; Metzinger 2005)), migraine, trauma,[11] general anesthesia (Bünning and Blanke 2005), near death experiences as well as neuropsychiatric disorders and neurologic disease. While there is much to be said about OBEs (see Table 3 for key features), I wish to emphasize only 3 points which will prove crucial for the later phenomenological analysis:

1. How is the "I" or putative first person perspective experienced in OBEs?
2. What is the role of movement (e.g., the respective visual, vestibular and/or proprioceptive contributions)? If there are abnormalities of bodily experience (including a possible separation of body schema (the "I") from body image (the "me"), to what extent is visual space reorganized in terms of the respective egocentric and allocentric reference frames?
3. Although patients and healthy subjects often report "lucidity" during these experiences, it is not clear to what extent the subject's own state of consciousness or background emotional-motivational state (as in Type II autoscopy) may play in how the subject constructs or "constitutes" (Husserl) the experience.

Contrary to autoscopy, OBEs are often felt to be positive (Twemlow et al. 1982) (e.g., a feeling of "freedom" or "control" in sleep paralysis). However, when precursor to a seizure (e.g., Case 5), they are experienced negatively. Sometimes, OBEs

---

[10] Wisdom (1953) distinguishes his concept of "phantom body" from body image: "Thus when one sees oneself in a dream or in a mirror the part of the visual body-image that is seen does not coincide with the phantom-body … That is to say that no fusion takes place between the phenomena these refer to: for in fact we usually find an approximate coincidence between the phantom-body and visual body-image." That is, the two selves (i.e., perceptual and motor, or body image and body schema, respectively) are usually integrated in everyday experience but are vulnerable to separation in neuropsychiatric disorders and anomalous conscious states. For the strikingly similar views expressed by phenomenological "anthropological" thinkers Plessner and von Weizsäcker, see Mishara (in press c): "The constant shifting between being/having a body – what Pleesner eccentric positionality – is often "overlooked" in everyday common-sense experience (von Weizsäcker; Plessner)" but disrupted in neurologic or psychiatric disorder.

[11] For example, rape victims sometimes report dissociative experiences, "such as the sense of watching the rape from outside their bodies" (Foa and Riggs 1993, p. 281).

Table 3 Key features of out of body experiences (OBEs) and the feeling of a presence (FOP)

Out of Body Experiences (OBEs)

- Feeling of separation from own (visible) body which – although not a requirement – may be observed from an external viewpoint
- Detached location or bodiless point of view is usually elevated and sometimes confined to right hemifield. When seen, one's own body is usually lying down or supine (Blanke and Mohr 2005)
- Current perspective is experienced as distinct "point of space" or within an imaginary 'second body' (a "phantom body," Wisdom 1953) which may be described as replica of one's own body (Brugger et al. 1997)
- OBEs are reported in 10% of healthy individuals (Irwin 1985)
- OBE experience often felt to be of great emotional significance for the subject; contrary to autoscopy, oten felt to be positive (Twemlow et al. 1982) (e.g., a feeling of "freedom" or "control" in sleep paralysis but may be a very negative experience when precursor to seizure, as in case 5)
- Unlike mirror autoscopy, the body does not look transparent, may cast a shadow (Devinsky et al. 1989 case 4)
- Mirror reversal (of spatial coordinates), when present, is not of one's own observed body but of the visually experienced environment (Brugger 2002; Green 1968; Bruce 1999)
- There is no ambiguity about location of self which remains with the first person perspective which is now bodiless or maintained by a "phantom body" (Brugger 2002; Wisdom 1953)
- Multimodel: may be accompanied by buzzing, snapping or ringing noises (Jantz and Beringer 1944) and/or vestibular sensations

The Feeling of a Presence (FOP)

- Conceived as an awareness of one's invisible *Doppelgänger*
- Despite the lack of any visual information, the spatial localization of the 'presence' is usually experienced at a precise distance from the subject (Brugger et al. 1996)
- Presence is very convincing, "intensely more real than any ordinary perception" (James 1961; Brugger et al. 1996)
- When a lesion is responsible, the presence is often confined to hemispace contralteral to lesion (Brugger et al. 1996)

Relation of FOP to OBEs, Type II Autoscopy and Psychotic Symptoms

- FOP may indicate a transitional step in the development of acute psychotic symptoms (e.g., auditory hallucinations, delusions) (Jaspers 1946)
- In sleep paralysis, FOP is associated with terrifying, nightmarish hallucinations, the feeling of vulnerability and – being temporarily paralyzed – the inability to defend oneself (Cheyne 2003, 2005)
- In sleep paralysis, FOP is more frequently associated with the hallucination of an incubus (and the accompanying feelings of suffocation, bodily pain and pressure on the chest) than with the more "blissful experience" of OBEs (Cheyne 2003, 2005)
- The increased occurrence of OBEs over time in the experienced sleep paralysis patient suggests that the patient has learned to convert a terrifying passive experience into one in which the patient regains some sense (even if illusory) of control and mobility
- The fact that FOP and OBEs can occur in the same disorder (e.g., sleep paralysis) but usually not in the same episode suggests that related but non-identical neural mechanisms may be implicated
- Type II Autoscopy may transition to FOP (Brugger et al. 2006)

are dominated by a sense of passively floating (Metzinger 2005, p. 67).[12] Interestingly, the fact that both OBEs and the feeling of a presence are associated with the hypnagogic/hypnopompic imagery of sleep paralysis suggests that their underlying cognitive-neural, mechanisms, although different, as I, shall argue may be related.

Feeling of a Shadowy Presence (FOP): I am Confronted by Another "I"

Case 6

   "The patient Kr. (dementia praecox (i.e., schizophrenia)) reports, 'I had the feeling that somebody was inside me and then, how would you say it, left my side? It was such a strange feeling. Afterwards, the 'someone' *always walked alongside me.* If I stood up, he stood up. If I started to walk, he started to walk. He always remained at the same place [behind me]. If I turned around to see him, *he also turned around at the same time* so that I was unable to see him. …Sometimes, I have the feeling that he comes closer or steps back a bit …' – The patient has never touched nor

---

[12] In his short story, "The Bucket Rider," Kafka (1979) depicts the feeling of a passively floating OBE-like experience humorously: The protagonist has the unique ability to ride his bucket through the village precisely because he has run out of coal and the bucket is empty: "I must have coal; I cannot freeze to death; … Seated on my bucket, my hands on the handle, the simplest kind of bridle, I propel myself with difficulty down the stairs, but once downstairs my bucket ascends superbly, superbly" (pp. 412–413). While hovering outside the coal-dealer's window and appealing to the latter's wife for some coal, he describes his experience: "She sees and hears nothing; but all the same she loosens her apron strings and waves her apron to waft me away … My bucket has all the virtues of a good steed except powers of resistance, which it has not; it is too light; a woman's apron can make it fly through the air … And with that I ascend into the regions of the ice mountains and am lost forever." (p. 414). Kafka frequently wrote at night during a sleep-deprived state. It is likely that he drew from his own hypnagogic imagery for this and other stories. In his *Diaries*, Kafka describes his nocturnal writing as conducted "entirely in darkness, deep in his workshop" (Kafka 1965, p. 518; see also Kurz, 1980). After writing "The Judgment" in one sitting, he praises the advantages of writing without sleep. It enables access to unusual thoughts and associations which otherwise would not be possible: "How easily everything can be said as if a great fire had been prepared for all these things in which the strangest thoughts emerge and again disappear" (Kafka 1965, pp. 293–294, my translation). That is, during writing, he experiences a transformed state of consciousness: "All I possess are certain powers which at a depth almost inaccessible at normal conditions, shape themselves into literature …" (*Letters to Felice*, 1973, p. 270)." Perhaps reflecting these abnormal, hypnogogic-like states, there are numerous examples in Kafka's (1979) writings of autoscopic doubling in which the mirroring becomes obstructive to the protagonist's own objectives. In "Descriptions of a Struggle" (Kafka's earliest published story, version A written 1903–1904; version B, 1909), the narrator's companion (an acquaintance just made at a party) continually does the opposite to what the protagonist anticipates, e.g., he walks too slow or too fast. However (as in Type II autoscopy), the narrator suddenly finds himself so embroiled that he catches himself mirroring (!) his companion: "he began walking again and I followed without realizing it…" (Kafka, 1973, p. 13) As much as the narrator desires to escape, he is unable to disentangle himself from the acquaintance as if they were each incomplete, or different sides of the same person. The "literary evidence" supports the phenomenological argument explicated here that autoscopy and intersubjectivity are related. For the phenomenological analysis of how Kafka's writings may reflect a hypnagogic process of hallucinating paranoid doubles of self (following social, sensory and sleep deprivation) with reference to Husserl's theory of intersubjectivity and the possible underlying cortical excitability of a social network in the brain, see Mishara (in press d).

directly perceived the presence. Nevertheless, he feels himself observed (by the presence), and insists that the whole experience is not an illusion." (Jaspers 1913, p. 415, my translation and parenthetical insertions, emphases in original).

In the feeling of a presence (FOP), another subject is *felt* to be in the self's proximity but not seen. It has been variously called "hallucination du compagnon" (Lhermitte 1939), "awareness of a (non-perceived) embodied presence" (*leibhafte Bewusstheit*) (Jaspers 1913), and "heautoscopy without an optical image" (Menninger-Lerchenthal 1935). As the previously described autoscopy-related experiences, it is associated with neuropsychiatric disorders. However, individuals, who have been isolated for long periods of time (e.g., mountaineers, explorers, sailors, and castaways), also report FOP.[13] Although the alien presence is not perceptual, it is often felt to be at a *precise distance* from the subject.[14] The presence is persuasive about its reality. Even when the patient acknowledges that the presence is not "real," he still expresses an insuppressible urge to offer the presence something to eat or a chair to sit on (Brugger et al. 1996).

The phenomenological-psychiatrist, Jaspers (1913, 1946) noted that it is not possible to classify FOP (*leibhafte Bewusstheit,* e.g., Case 6)[15] as a perceptual hallucination or as a delusional belief. Nevertheless, it may harbinger a subsequent "transition" to hallucinations and/or delusions (Jaspers 1946, p. 67). FOP may serve as a stage in the development of these other psychopathological phenomena and continue to be a component of these experiences. For example, prior to the development of more florid symptoms of active paranoid psychosis, prodromal schizophrenia patients may feel "they are 'watched' or 'observed' without anyone nearby" (Jaspers 1946, p. 67, my translation). Although the idea – as far as I know – has not received attention in the current literature, the phenomenology and underlying neurobiological mechanisms of FOP may indeed support the view that FOP is a critical component in the development of delusions and hallucinations in schizophrenia.[16]

---

[13] This concurs with Ralph Hoffman's (2007) hypothesis that sensory/social deafferentation (and associated cortical excitability) in patients with schizophrenia contribute to their positive symptoms (hallucinations, delusions). Similarly, subjects who participated in John Lilly's "isolation" (sensory-deprivation) tank experiments (floating in darkness with the salt-water at body temperature) reported both hallucinations, and eventually, delusions.

[14] "Several observations make clear that this 'person,' often referred to as a shadow, is nothing more than a projection of the own body representation into extracorporeal space" (Brugger, 2007, p. 212).

[15] Jaspers (1913) draws from Ach's (1905) term *Bewusstheit,* i.e., the making present of something known to 'awareness' but not directly given to perception (*Bewusstheit* = *Gegenwärtigsein eines unanschaulich gegebenen Wissens* (Ach 1905)).

[16] Hoffman et al. (2003) found that slow repetitive TMS, which decreases the excitability of the underlying cortex, reduced the incidence and severity of treatment-resistant auditory hallucinations (when applied to the left temporo-parietal region). However, even after alleviation from the distressing hallucinations, some patients report that they can still "feel the voices," i.e., the "presence" that gives rise to them (R. Hoffman, personal communication). Presumably, these feelings arise from activity in neighboring brain areas (see Fig. 2 and discussion of neural correlates, below). As further indication of a presence 'behind' the voices, I interviewed a schizophrenia patient who refers to her recurrent auditory hallucinations as "presences." She reported that the "presences will wait to talk" with her while she is doing some other activity such as reading. However, she senses their "presence" so strongly that she feels obliged to stop reading and initiate "conversation" with them.

Right-hemispheric epilepsy patients (Kamiya and Okamoto 1982) sometimes report "mental dipoplia" (i.e., the feeling of mental duality, Hughlings-Jackson 1932), which resembles FOP. The invisible double may be experienced outside or inside the body and may speak with the subject from this vantage point. For example, one patient heard the voice of his other self from his abdomen (Kamiya and Okamoto 1982, Case 6). The patient may also report a "double thinking," the sensation that two selves are thinking within the same subject.

Both FOP and OBEs occur in sleep paralysis but rarely in the same episode. In sleep paralysis, the patient experiences an inability to move when falling asleep or upon waking. Although conscious of the surroundings, the patient may, at the same time, have nightmarish hallucinatory experiences. Cheyne (2003, 2005) classifies these experiences into three kinds: "Intruder, Incubus and Vestibular-Motor."

Intruder-hallucinations (the equivalent of FOP) involve experiencing a numinous presence (without direct perception), usually sensed to be evil or threatening. It may be accompanied, or followed by multimodal hallucinations, including vague rustling sounds, indistinct voices, demonic gibberish, "sensations of being touched or grabbed," and/or visual hallucinations of human apparitions, animals or supernatural creatures (Cheyne 2003, p. 164). Cheyne (2003) states that "the role of the other" is an important construct for understanding Intruder-hallucinations (p. 175), an observation which will be supported by the phenomenological analysis.

Incubus-hallucinations involve feelings of suffocation, breathing difficulty, bodily pain and pressure on the chest, sometimes interpreted as resulting from some entity climbing onto the chest of the patient (depicted in Fuselli's well-known, 1781 painting, "The Nightmare"). Intruder- and incubus-hallucinations are reported to be distressing, even terrifying, constituting a "waking nightmare," and may co-occur in the same sleep paralysis episode.

Due to sensations of bodily acceleration, including flying and floating, the third type of hallucination, vestibular-motor, is often experienced as an OBE. In contrast to intruder- and incubus-hallucinations, these experiences are associated with "blissful feelings." They increase as the patient becomes more experienced and are less likely to co-occur with the two other hallucination-types in the same episode. This suggests that the patient has learned to convert a terrifying *passive* experience (e.g., FOP) into one in which the patient regains some sense (even if only illusory) of *control and mobility*. In a disorder in which the primary symptom is motor paralysis, the fact that patients are able to transform an oppressive experience into a more positive one suggests that *the experience (and its perceptual and motoric features) becomes reorganized*. Critically, the fact that FOP and OBEs may occur in the same disorder (e.g., sleep paralysis) but not in the same episode suggests that related but different neural mechanisms may be implicated. That is, the feelings of *vulnerability vs. mobility* of body, *active vs. passive* attitudes, *self vs. other* appear to be reversed in FOP and OBE but the precise relationships and the underlying neurobiological mechanisms need to be clarified.

As FOP and OBE may be related (albeit in mutually exclusive relationship) in the same disorder, some case reports suggest that autoscopy and FOP may also be related:

Case 7

A 41-year-old, male pottery maker, PH, with an invasive tumor originating in left posterior insula (destructive of left-temporal lobe, extending into frontal and parietal areas), awakens one night to notice that he has split into three distinct parts: (1) the left-half of his body felt quite normal; (2) the right-half felt detached from the left both physically and emotionally; and (3) a man in close proximity to his right side which he felt to be a part of himself, "sharing the same soul" (i.e., Type II autoscopy). However, there was no similarity in physical appearance (the man was blond, while the patient's hair was black): "…*trying to get a close look of him, I all of a sudden noticed that, even more to the right, there was a whole group of people. At a distance of 2 m I saw an approximately 50-year-old lady with blond braids. Still another 4 m away, there were two girls* (both approximately age 20) *and some 20 m from me, still in a straight line with all the other persons, there was a boy* (unspecified age)… *Naturally, I could not see the persons any longer on closing my eyes, but the feeling remained that pieces of myself were located in precisely those places I knew the persons were standing"* (See Fig. 1). Although this family did not resemble his own family (the patient's real wife was younger than him and had dark short hair; his only two children were two sons, aged 10 and 16), all members were felt to be related to himself. *Moreover, the family members imitated PH's every movement.* The exception were the two girls who, more distal spatially and "talking to one another, would look towards me waving their hands as if inviting me to join their world." When PH's actual wife sat by his right side, the hallucinated family would temporarily vanish but he then noted a clumsiness and weakness to his right side. After 2 h, the patient manages to fall back to sleep.

Next morning, the patient is taken to the hospital. At this point, the patient no longer sees the family but *feels their presence* (FOP): "'some hardly describable sense' made him aware that the 'family' was still present and enabled him to precisely localize the position of four persons in his room. Specifically, he noticed that the 'father' had moved to the right [i.e., no longer on the symptomatic side] while the distance to the 'girls' had shrunk, such that the "family" now gathered … 2–3 m from his side with the exception of the "son", who had disappeared. … The patient felt that each member of the 'family' was equally a 'part of (his) expanded self'." *They continued to mirror his movements.* Later the same day, they began to communicate with him (by transferring their thoughts to him rather than by 'normal' means of verbal communication). Upon awakening the 3rd day, the sensed family (FOP) disappeared (Brugger et al. 2006, in part modified, pp. 669–670).

Several points concerning this remarkable case are relevant to the subsequent phenomenological analysis: (a) The hallucination had dreamlike qualities. For example, although the imagined persons did not resemble himself and his family, PH "knew" (all at once) that there was a connection with himself; (b) When his actual wife sat to his right side, the figures would disappear only to be replaced by an increased awareness of symptoms (i.e.," clumsiness and weakness"); (c) When he closed his eyes, he was still able to sense the places where the presences stood;
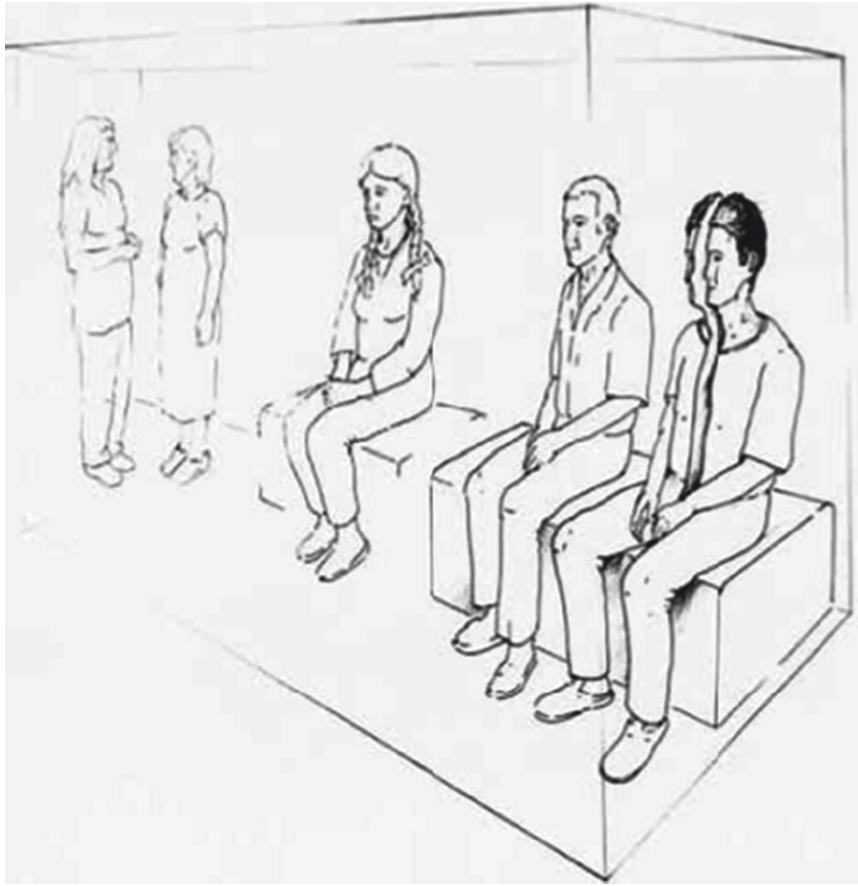
Fig. 1 Artist's drawing according to PH's verbal description of four of his five doubles. All doubles were invariably located to the right of the patient's body. Despite nobody of the "family" of doubles reflected the patients' appearance, all members were felt to belong to the own self. The degree of motor and psychological autonomy of a double increased with increasing distance from PH's body. The presence of one additional figure was vaguely perceived (more 'sensed' than seen) at a distance of about 20 m from the patient (not displayed). The space toward the patient's right side appeared to be stretched (Reprinted from *Cortex,* 42, Brugger, P., Blanke, O., Regard, M., Bradford, D.T., and Landis, T. Polyopic heautoscopy: Case report and review of the literature, Copyright (2006), 666–674, with permission from Elsevier)

(d) After the visual apparitions had disappeared, there was a residual period in which their presence and continuously imitating behaviors (now more closely surrounding the patient) were only sensed but not seen. Points (c) and (d) suggest that Type II autoscopy (in this case polyopic) and FOP co-occur in the same disorder.[17] The closing eyes experiment (c) suggests that FOP may be an inherent component of Type II autoscopy and the two may be related in terms of underlying mechanisms.

---

[17] Peter Brugger reports having observed other (unpublished) cases in which an initially visual experience of heautoscopy (Type II autoscopy) recedes over days or even weeks to later become the *feeling* of an invisible double (FOP) (personal communication).

The phenomenological-neurologist, von Weizsäcker (1948), labeled relationships of functional reorganization between symptoms and awareness (e.g., (b)) or between symptoms (d), "substitution by means of a representative" (*Stellvertretung*). Each expression of illness (b, d) replaces the other but thereby induces a *reorganization* of the subject's experience in order to maintain a "coherent" self-world relationship. Since many readers may not be familiar with von Weizsäcker's concepts, I will present them in further detail below.

Phenomenological Approaches to Autoscopy

Phenomenological method (as proposed by Husserl) offers a disciplined sequence of steps (Mishara and Schwartz 1997): (1) Phenomenological reduction is a "leading back" (from the Latin *reducere*) from one's current engagement with the world to examine (reflectively) the "streaming-consciousness" in the here and now; this requires the bracketing of common sense folk-psychological/folk-physical assumptions about how minds and objects behave in the world; (2) Abstracting the essential meaning-structure of an object by bracketing (or suspending) its reference to reality and examining its limits by freely imagining variants which fall within its semantic boundaries ("eidetic, imaginative variation");[18] (3) Rigorously describing the "results" in a technical language which is as sensitive as possible to fine details of the experiencing while monitoring this language for lapses into reification of our common sense folk-psychology; (4) Integrating the findings into a theoretical

---

[18] Husserl's method of abstracting, or delimiting the meaning of an object by means of imaginative variation of its limits, establishes the formal conditions of its possibility (see Seebohm 1962). By suspending our belief in the object's *actual* reality (Step 1), we explore its *possible* imagined, variant meanings (latent or implicit to the initial perception) (Step 2). What is common to each variant (i.e., a sample of the core-meaning according to the underlying "type") is actively maintained across variants. After sufficient sampling, there results a "seeing of the object's essential structure" (*Wesenschau*) initially present (in a concealed manner) in the perception (see the discussion of perceptual "type" below). By means of imagination alone, Goethe (1790) thought that he accessed a "primal plant" (*Urpflanze*) that provides the "formula, rule, or law" (*Formel, Regel, Gesetz*) for the appearance of plants in general. This approach may be considered a distant precedent to Husserl's method. Unlike Husserl, however, who was proposing a method to discover the underlying structures of consciousness, Goethe thought that he was directly studying the laws of nature. C.G. Jung (a contemporary, whom we know Husserl had read from underlinings in his library copy of Jung's *Versuch einer Darstellung der Psychoanalytischen Theorie*, 1913) had developed a related method called "active imagination." This method is intended to reveal unconscious meaningful structures (so-called "archetypes"), which (not unlike Husserl's type or eidos), once triggered or activated, play a role in 'determining' the conscious experience. Phenomenological-psychiatrists and researchers (e.g., Binswanger, Buytendijk, Tellenbach) justify their use of literature as a source for phenomenological "data" for providing the structures of healthy (and abnormal) consciousness by referring to Husserl's second methodical step, imaginative variation.

framework which is then imparted to a community of investigators for "replication" using the same method.[19]

As method of distancing from our current assumptions, phenomenology "brackets" the presuppositions of common sense that inform our folk-psychological views to study the otherwise nonapparent, concealed structures of consciousness (Step 1, above). It provides a different starting point for the analysis of disrupted self-experience in neuropsychiatric disorders than the apparent (usually unquestioned) "givens" of our common sense.[20] We have to be particularly vigilant about how common sense distinctions uncritically inform our explanatory constructs or reifications about mind when thinking about problems of cognition.

Apart from Conrad (1953) and Jaspers (1914), Zutt (1953) is one of the few phenomenological-psychiatrists to discuss autoscopy-related experiences. He remarks that one reason why we have so much difficulty understanding the underlying structure of OBEs is that we begin with *the common sense assumption that we ordinarily experience the self, the "I," as encapsulated "in the body." By reflecting on our experience, we place the self back "into the body" not, however, as an I but as an object or "me."* In fact, Husserl (1966) emphasizes that thinking about or reflecting on our own "current" experience is by necessity retrospective.[21] The very act of thinking about or reflecting imposes *a splitting of the "I" (Ichspaltung) into an currently thinking or reflecting "I" and a reflected (already past !) "me,"* whereby, there is no evidence – being very careful in our descriptions – that these terms were actually separate in the original experience. [22] As Husserl

---

[19] The phenomenological method (which I present here in a simplified manner) has been criticized as being less transparent than it claims (e.g., Gadamer 1993). It is very difficult to explicitly follow a method without implicitly employing the eventual "expertise" that accrues through practice (i.e., the procedural know-how and accompanying "prejudices" (*Vor-urteile*) we acquire over time and apply without awareness) (See Mishara, 2007b). Similarly, Merleau-Ponty (1973) observes that phenomenological reduction (i.e., reflection on our experience) is inevitably mediated by language even when we claim to be describing nonverbal, pre-linguistic, "mute" sources of meaning (see the discussion of "type" below).

[20] The phenomenological-psychiatrist, "Blankenburg and Mishara ((1969) 2001; 1971) observes that "our mental health is preserved by a certain 'resistance' to loosing our common sense. This resistance functions precisely by overlooking the obvious as obvious. The obvious … does not require further exploration; it even 'resists' further exploration" (Mishara 2001, pp. 319–320).

[21] Every reflecting on our experience itself occurs within the temporal passing of consciousness and is subject to the same "laws" of "time consciousness" as the original reflected on experience (see Yamaguchi 1982).

[22] These observations, of course, do not overlook "prospective" reflection (see Buckner and Caroll 2007, for recent review). However, in both prospective and retrospective (explicit) *reflecting* – even as I *plan* tomorrow's activities – the self is experienced *as object.* The episodic/declarative memory system (retrospective or prospective) is mediated by medial temporal lobe, which receives input from both ventral and dorsal visual processing streams, so-called "what" (allocentric) and "where" (egocentric) pathways, respectively (Eichenbaum et al. 2007). Nevertheless, *explicit* remembering and prospection may be exclusively cast in scene-based, allocentric coordinates, in which the "remembered" or "projected" self is always object or a me. It may be impossible to have a reflective or explicit view of self which is not object (Mishara 2007b). The first person spoken "I" (see below) does nothing to change the reflected self's status as object.

(1966) observes, there is a forgetting or loss of self (*Selbstverlorenheit*) in naïve experiencing.[23] To reflect on the self is to recover it from this self-forgetting but in a manner which (inevitably) reflection itself imposes. Zutt raises the question "where" in fact is the "I" during our original prereflective experience of things? When I perceive something, I unreflectively become engaged with it. He writes, "when I engage with something out there in space, I am there with the thing." (p. 26 my translation).[24]

Now, when we explicitly reflect on ourselves experiencing something, the self momentarily returns to its default location (possibly artifactual to reflection itself), "inside the body" or (if you will) "behind the eyes." Because contemporary neuroscience tends to locate the "first-person perspective" within the body, many researchers regard OBEs and autoscopy in terms of a "displacement" of self from its natural location "in the body." For example, Ehrrson (2007) concludes that OBEs disrupt the sense of self as being "located inside the body": "This finding represents a fundamental advance because the natural 'in-body experience' forms the foundation for self-consciousness" (p. 1048). This conclusion, however, is based on an incorrect understanding (from a phenomenological point of view) of the embodied self as being-in-the-world (see phenomenological arguments below).

The phenomenological approach differs from current clinical neuroscience efforts to explain autoscopy. For example, Lopez et al. (2008) propose two related cognitive-neural mechanisms: (1) self-attribution = body ownership; (2) Self-localization = embodiment. Both mechanisms are disturbed in Type II autoscopy and OBEs due to impaired vestibular and somatosensory (proprioceptive) information processing. In FOP, however, embodiment and body ownership are dissociated as the feeling-of-a-presence is characterized by disturbed body ownership (or self-attribution of the illusory body) but normal embodiment. In this condition disturbed body ownership has been linked to abnormal sensorimotor processing (rather than

---

[23] Husserl (1966) observes that, when we are caught up in perceiving or experiencing things, there is a loss of self (*Selbstverlorenheit*), a naivety about our role in constructing the experience: "Admittedly, the moment I begin to reflect, the naïve perceiving by the self-forgetting I is already past. I am only able to grasp this by reaching back – in the reflecting – into what has 'remained in consciousness' as retention, an immediate memory which *attaches itself backwards* to the original experience" (p. 88, my translation and emphases). I am able to reflect on my original naïve self-forgetting which is absorbed in the experiencing only because the I itself has 'split' (*Ichspaltung*) into a reflecting I and the object of its reflection, the naïve I just previously engrossed in experiencing (i.e., the self as object or a "me"). Husserl argues that we are nevertheless assured that they are the same "I" because both are experienced in the retrospective reflection as "belonging to the same streaming present" (p. 89, my translation).

[24] In this regard, Zutt (1953, p. 26) cites the phenomenological-philosopher, Heidegger (1927), "In directing one's attention to … or grasping something, the existing subject (*Dasein*) does not proceed from some 'inner sphere' in which he is at first encapsulated. Rather, the subject's primary mode of being is to be always already 'out there' with the things he is encountering, with the world as it discloses itself" (1927, p. 62, my translation). That is, the being-in-the-world of the subject is ec-static (from the ancient Greek, *ek* out + *histanai,* cause to stand), i.e., as a subject standing outside him- or herself in engagement with the world.

disturbed vestibular and proprioceptive processing). In Type I autoscopic hallucinations, both impairments are absent or milder.

In a similar vein, Blanke et al. (2004) and Blanke and Mohr (2005) suggest that autoscopic phenomena result from a failure to integrate multisensory bodily information. These authors propose that autoscopic phenomena result from a disintegration in bodily, personal space (due to conflicting tactile, proprioceptive, kinesthetic, and visual information) and a second disintegration between personal and extrapersonal space (due to conflicting visual and vestibular information).

In contrast to these views, the phenomenological approach offers an alternative set of hypotheses: The embodied relation self and other(s) (i.e., intersubjectivity) becomes disrupted in autoscopic experiences; This involves a breakdown in preattentive "efferent binding" (Haggard et al. 2002) between subcomponents of self (e.g., "I," "me" and "mine," so far as these are captured in language) prior to the emergence of self as a unitary experience in awareness (Mishara 2007a).

The Other Is Like Me, I Am Like the Other(s)

Philosophical approaches (phenomenological, analytic) and social cognitive neuroscience are confronted by the problem of how it is that we experience other persons at first as perceptual objects (in German, *Koerper*). Nevertheless, we attribute to them minds and experiences, i.e., subjective experiences, not directly available to inspection. In Husserl's (1959) view, the experience of *my* "body as subject" (in German, *Leib*, related to *leben*, to live) plays a fundamental role in the experience of others as embodied subjects. It serves as the prototypical body subject (*Urleib*) for how I experience others as also embodied.[25] On the other hand, Husserl (1988) writes: "I do not apperceive the other ego simply as a duplicate of myself … rather, … I apperceive him as having spatial modes of appearance like those I should have if I should go over there and be where he is" (p. 117).[26] My being a body subject is immediately given as my own. The other's internal (mental) relation to body is never given originally but it is experienced (in empathy) as *analogous* to my own. Thus, the other's body is "not a second body subject belonging to me" (Husserl 1988, p. 113).[27]

---

[25] "This is so because my body is already always there in the perceptual field as body subject…" (Husserl 1959, p. 62, my translation).

[26] When perceiving others, what is at first "… given is the outside of another living body. This is apprehended as **living** because an inside is associated with it through an associated transfer of the **inside/outside structure** given with my primordial body. This inside appresents (i.e., makes present) another simultaneous here-now which is not mine and cannot be united with my own here-now …" (Seebohm 1989, p. 373, my parenthetical insertion and emphases).

[27] Because the other's body is always "there" and can never occupy the "absolute Here" of my body, "the other body can never be given to me as my own living-body" (Held 2003, p. 52).

Without direct access to the other's inner experience, nor he to mine, we find ourselves in an inextricable reciprocal relationship of assumptions about our respective embodied "minds." "I cannot help putting the other, and the perception he has, *behind his body*." (Merleau-Ponty 1968, p. 9). By a sort of identification with his perspective, I also locate my own "self" *behind* (or in) my body as if experienced from an external viewpoint.

To summarize:

1. I experience the other as experiencing a here and now just as I do from my perspective. The other's presence awakens in me the experience of his perspective "as if I were there" (CM 147) in the place where the other is. This occurs in a passive, automatic or obligatory manner (i.e., not as conscious inference). Our two perspectives are experienced as mutually exclusive, repelling one another without danger of fusion.
2. It is true that I am also able to experience my own body as object (*Koerper*) and I presume that others have the same inner relationship to their bodies as I do to mine but I have no access to this presumed operation in the other.
3. Moreover, the "two levels of bodily experience" (inner-mental, outer-object) in myself are *perceived* as one (Husserl 1959, p. 61). For example, when I move my hand to grasp an object or kinesthetically orient my eyes to look at something, I experience these as one "act" (i.e., perceptual and motor processes are preattentively integrated in a perception action cycle, von Weizsäcker 1950) or the "efferent binding" of perception, agency and self (Haggard et al. 2002).

In Type II autoscopy, the mutual exclusion of perspectives (Case 1) and exclusive access to my own mind (Case 2) are disrupted. In OBEs, the experience of my body as subject is separated from the experience of my body as object (Case 3). Up to this point, the phenomenological analysis and autoscopy appear to be in direct contradiction.

If one continues this analysis on "deeper," more automatic levels of functioning, however, a different picture emerges.[28] Becoming conscious of the self's and others' bodies occurs in a "reciprocal awakening (*wechselseitiges Sich-wecken*)." "The reciprocal effect is not one-sided from my own activity, but develops from a purely passive associative synthesis which is not to be labeled as 'my own'." (Yamaguchi 1982, p. 96, my translation). In this sense, there is a subpersonal, "anonymously functioning embodiment" – what Merleau-Ponty (1968), metaphorically calls

---

[28] Husserl's phenomenology anticipated later developments in cognitive science. For example, Husserl's distinction between "active" and "passive" mental processes overlaps with the later cognitive distinction between controlled (i.e., effortful, limited capacity) vs. automatic processing (Mishara, 1990; Wiggins, 1994; Wiggins and Spitzer, 1997). For Husserl's account of how phenomenology penetrates into "deeper" layers of nonconscious, automatic processes, see Mishara, 1990. Similarly, Binswanger (1957; 1965) regards psychosis as a "natural experiment" in which the researcher is able to examine "deeper," nonconscious levels of meaning. It is as if the layers of our experience were suddenly exposed, layers that otherwise would not be accessible (Mishara in press a).

"flesh" [29] – which does not prioritize self or other, but finds them *in a precisely calibrated dance* between inner and outer, and – as we shall see – active and passive, movement and perception.

The phenomenological theory presented here proposes (a) that *autoscopy is a breakdown in intersubjectivity*, a reciprocity between self and other(s) as "alike," "mirroring" one another (Husserl 1960, p. 94); (b) different forms of breakdown are exhibited in the types of autoscopy; (c) the descriptive-clinical and neurobiological findings regarding autoscopy can be integrated with one another.

Mirroring as Self-alienation

When considered at a "deep" enough level of automatic processing or passive synthesis (Mishara 1990; Yamaguchi 1982), the relationship of likeness between self and other, as we have seen, emerges without priority. On "higher" levels of processing, however, the following conundrum occurs: At the same time that I have exclusive access to my own inner experience of body, the other has privileged access to my outer, visible body which he grasps as a unitary, mobile Gestalt. He does so in a way (given my limited, partial view of my own body) that I am unable. Conversely – although I am excluded access to his (inner) relationship to body as subject – I do have the same privilege of seeing his body externally as a totality (in a way that he is unable). As the French psychoanalyst, Lacan (1977) noted, the mirror allows the subject to envision the body-self from an external perspective as an "imaginary" unity, as how another *would* see the subject from outside as a moving visual totality. In this sense, the mirror is "profoundly alienating" (Stawarska 2004, p. 176). It accomplishes a split between "I" (what we have been calling body subject or *Leib*) and "me" (body as object or *Koerper*) which is mediated by the other's privileged perspective on my body as a totality (and my own ability to *imagine* this perspective through the mirror).

The mirror image (as my external double) is *a structure which my own self provides*. Nevertheless, it exercises a "*captivating*" fascination (not unlike the various forms

---

[29] Merleau-Ponty's project may be seen as a continuation of Husserl's radical approach to cognition. By revolutionizing our view of embodied cognition, he proposes a new vocabulary for the study of human self and experience: "Replace the notions of concept, idea, mind, representation with the notion of dimensions, articulation, level, hinges…" (1968, p. 224). In order to realize this project, Merleau-Ponty (1968) employs metaphorical language, e.g., "chiasm," "hinge," "flesh." These are meant to describe the relationship between a body for self (body subject) and body for others (body as object) as an ongoing, "reversible" relationship (critical for the current phenomenological approach to autoscopy). Many of the passages that employ these metaphors, which I cite here, are written in note form for a never completed book project (1968) shortly before Merleau-Ponty's untimely death. Therefore, they are suggestive but, unfortunately, incomplete. See Mishara (2007a) for the (inevitable) role metaphors play in both the self's own expression and its neuroscientific study.

of autoscopy we have been describing). The mirror fascinates because I imagine how I appear (passively) as body-object to others. In his analysis of children's play, Merleau-Ponty (1964) observes that when a slightly older child "parades" before a younger child, taking an active, "despotic" role (i.e., "playing with this or that latest toy, talking, holding forth"), there follows a similar "captivation" of the younger child's attention. The older child's active, dominant role "provokes in the [younger] child the *complementary* (passive) attitude" (p. 142, my parenthetical insertions and emphases).

Such automatic and obligatory complementary attitudes are also present during conversational turn-taking. That is, when I listen to the other's speech, I take on an outwardly passive attitude. At the same time, I (actively) anticipate and reproduce his spoken meanings to myself. For the moment, I am the other but in a way hidden to myself.[30]

Turn-Taking with Others and with Myself

In the phenomenological theory of object-perception, a *type* (Husserl 1973) or *schema* (Binswanger 1965) provides the principles of organization, i.e., rules for (passive) synthesis of aspects of the perceptual-object in its inner structure. Because the type is anticipatory, it provides the rules by which each partial view is in turn synthesized into an object, a totality that is never apprehended by just one view.[31] Just as I do not currently perceive but anticipate an "other side" to perceptual things, so I impute to another living being an "other side," i.e., subjective experience. However, unlike the perceptual object, this will never be directly experienced but given to me only "as absence" (Merleau-Ponty 1968, p. 168). That is, I and other(s) experience one another's (inner) subjectivity as denied access, mutual exclusion or

---

[30] Merleau-Ponty describes how one "structure" shared between conversational partners enables the (automatic, anticipatory) transitions required for turn-taking: "Speaking and listening, action and perception, are quite distinct operations for me only when I reflect on them.… The synthesis of coupling or *transition* … (of) looking/gesture, listening/speaking … begin as simple modalities of perception and movement … At two levels, the recognition of the passive by the active and of the active by the passive, of the person spoken to by the speaker, is projection and introjection … He is the person spoken to, i.e. an offshoot of myself, outside, my *double*, … because I make him do everything I do and he makes me do the same" (Merleau-Ponty 1973, pp. 19–20, my emphases).

[31] Object-perception requires the ongoing synthesis or binding of (1) the currently experienced aspect with (2) the aspects not available to current perception and (3) a totality or unity of aspects that is never actually given in terms of the one aspect. The type or schema implicitly organizes the aspects of an object into a coherent relationship of perceptual meaning prior to its conscious perception. The fact that the object type or schema is already activated at this level of object recognition and is responsible for the inner mutual coherence of aspects or views of the object (as variants of the object's core but invariant perceptual meaning) allows for the seamless transition from an object's prelexical identification in visual perception to its linguistic expression in conscious, explicit judgments. (However, conscious awareness of the object does not itself require this transition to verbal naming) (see Uhlhaas and Mishara 2007).

"absence." The other person's subjective mind is experienced as "inner" because I put the other's mind *behind his body* as I experience him (presumably) doing with me. When thinking about myself (as object of the reflection), I inevitably do the same to myself.[32] Lacan's analysis of the mirror (and Sartre's analysis of the "Look")[33] provide evidence that the perception of the other as subject (their inner perspective) is to simultaneously experience my body now from an outer perspective as object. *That is, when I experience the other as subject, I experience my body-self, for the moment, in a complementary manner, as reduced to an object from his perspective.*

With regard to this last point, Merleau-Ponty (1968) describes the (oft-cited thought experiment of the) "double-sensation," i.e., one of my hands touches the other hand (which, in turn, touches some third object in the world): "My left hand is always on the verge of touching my right hand touching the things, but I never reach coincidence; the coincidence eclipses at the moment of realization, and one of two things always occurs: either my right hand really passes over to the rank of touched, but then its hold on the world is interrupted; or it retains its hold on the world and I do not really touch *it* – my right hand touching, I palpate with my left hand it's outer covering." (1968, pp. 147–148). I may, at will, alternatively inhabit each of the two complementary attitudes. However, the moment I actively touch my right hand with my left hand, my right hand becomes its object (something touched) and loses its active status. The moment my right hand resumes its active role touching the object, I no longer experience its being touched.

From the analyses of embodied intersubjectivity, we may conclude: (1) I "anticipate" but never access the other's internal subjective experience of his or her body; (2) nevertheless, I experience, with regard to the constituted "absence" (non-availability) of the other's (subjective) perspective, a *"complementary"* objectifying of my own body in an automatic and obligatory manner. From Merleau-Ponty's thought experiment, we now see these same conclusions also apply to my relationship to myself. At the moment I assume the role of passively being touched, I am no longer the one touching. *By some prearranged, prior unity, however, – which is not available to my current awareness – I am able to transition back to the one who is actively touching.* That is, to enter into this relationship with myself as capable of being both body subject and object, *I must presuppose an "other side," an inner subjective*

---

[32] My experience of *mind as "inner"* may, in part, be the artifact of my simulation of the other's perspective combined with my own reflective thinking. That is, reflection reconstructs the self as a me or object located *in the body* in the same way that I experience others experiencing me. In each case, I take an external relationship to myself as "having" a body, or "having" a self.

[33] Sartre's (1966) celebrated analysis of the "Look" indicates a similar structural relationship between self and other to the mirror image (Stawarska, 2004). In Sartre's account, the peeping Tom, by peering through the keyhole, transforms the target's body into an "object" (without concern for the other's "inner" subjective bodily experience). As the peeping Tom hears footsteps coming up the stairs, the objectifying "look" he had just exercised is now reversed on him. With the prospect of being caught in the act, he anticipates how someone else might see him, an object of (self-conscious) shame in the eyes of the other.

perspective (just as I had attributed a mind or subjective inner side to the other). However, *I am, for the moment, unable to directly access* (or in more contemporary terms, 'cognitively penetrate') this other active side, when I identify with the moment of being touched.[34]

The attentive reader may already see the relevance for autoscopy, especially Type II autoscopy: to the extent that my double becomes empowered with my subjectivity, I become derealized. I become the object, the (ironic!) 'reflection' of my own self. Whatever *seeming prearranged, prior unity which allows me to transition back to the one seeing, the active one, is now no longer available and I become caught in my own structure of self as now other, as no longer I, but only a 'me'. In OBEs, it is the reverse, I am unable to transition from the active seeing self back to my body as object*, the body as seen, touched, etc. Because the putative neural correlates of Type II autoscopy and OBEs are lateral homologues (i.e., involve the same brain area or structure but on opposite sides of the brain), they lend support for the phenomenological theory of autoscopy as disrupted self-other relationship (see "*Do the Neuroanatomical Correlates of Autoscopy Support the Phenomenological Theory?*" below). This involves being 'stuck' in either a predominantly passive (Type II autoscopy) or active (OBEs) attitude. The neurologic-phenomenological concepts "body schema" and "body image" further clarify how self-other relationship may be disrupted in autoscopy.

Body Schema/Body Image: Reversible Reference Frames Mediate Self-other Relationship

In their groundbreaking article, Gallagher and Cole (1995) define *body schema* as "a system of preconscious, subpersonal, anonymous processes that play a dynamic role in governing posture" (p. 369). Below the level of self-referential intentionality, it, at the same time, supports intentionality in all its forms. Despite not requiring consciousness, it is tacitly keyed into the environment. In contrast, *body image* "consists of a complex set of intentional states – perceptions, mental representations, beliefs, and attitudes – in which the intentional object of such states is one's own body" (Gallagher and Cole 1995, p. 369).

Similarly, for Paillard (1999), the distinction body-image/body-schema is that between "a conscious awareness of one's own body" and "a nonconscious performance of the body": "Proprioceptive information is obviously necessary for updating the postural body frame (or schema), whereas exteroceptive multimodal information, mainly visual, underpins the central representation and percept of the body image

---

[34] Similarly for vision, "I cannot see myself in movement, witness my own movement." (Merleau-Ponty 1968, p. 254). For example, the orienting "kinaestheses," i.e., ocular motor movements, are not themselves conscious when I view the things the orienting makes possible (Claesges, 1964; Husserl 1997).

…" (pp. 197–198). The body-schema provides a "path structure", superimposed on a collection of separate points, in a vectorial map which defines in *egocentric* terms how awareness is able to shift from a current "here" to an anticipated but still not consciously known "there." Paillard (1999, 2005) acknowledges the overlap of his model with Milner and Goodale's (1995) proposal for a "vision-for-perception" ventral system which is more recently evolved (mediating awareness) and the more ancient (nonconscious) "vision-for-action" dorsal system. For Milner and Goodale, the dorsal stream projecting from primary visual cortex to the superior parietal lobes is a key component in an action pathway of visual processing which locates "where" a relevant stimulus might be in the periphery relative to current focal vision (Paillard 1991a,b); Paillard 1999; and Paillard 2005).[35] Information may be relayed to the dorsal or ventral pathways based on its peripheral or central location in the visual field. *Information from the peripheral visual field has faster access to the implicit body-centered computations of dorsal processing streams than the slower ventral pathways subserving conscious focal awareness*. A third interoceptive pathway mediates the inexorable "mineness" of my experience and forms its (emotional-motivational) background. It is, in principle, dissociable from the other components "I" (body-schema) and "me" (body-image) (De Preester 2007; Mishara 2004).

The transition and thus, seamless binding between body as subject (body-schema) and body as object (body image) occurs in the ongoing sensorimotor integration required for bodily awareness *as always already mediated by the efferent binding of an ongoing perception action cycle*. The movement from an implicit egocentric to an explicit allocentric frame of awareness in "efferent binding" involves reversing frames of reference, i.e., the body-subject and body-object are each the inverse of the other.[36] The "efferent binding" of "I move myself" (for Husserl, the core of self-transcendence in time) involves both conscious and non-conscious components of a perception-action cycle (or *Gestaltkreis*, von Weizsäcker (1950).

---

[35] Nowak and Bullier (1997) coined the term "fast brain" for the fronto-parietal connectivity of the dorsal pathways which, according to the Goodale Milner model, mediate implicit visuomotor control (as well as sensori-motor transformations from other sensory modalities necessary for this control). That is, information coming from the peripheral visual field "has access to fast, direct pathways that allow for faster onset times in dorsal stream areas." Moreover, we may conclude that the function of frontodorsal connectivity is the "monitoring of peripheral stimuli in general" (Stephen et al. 2002, p. 3072). Remarkably, such a system of self as prospective openness, i.e., the ability to be affected by any point in its experiential field (structured by momentary, possible movement) prior to focal awareness had been anticipated by Husserl (Mishara 2005). The location in the field is *prospectively* structured by the "kinaestheses" of ocular motor response, i.e., by a potential field that is structured (nonconsciously) in terms of possible movements (i.e., eye-centered coordinates) required to reorient optimally to the novel target (Claesges 1964; Husserl 1997; Mishara 2005).

[36] Merleau-Ponty (1968) writes: By means of "reversibility … alone, there is passage from the 'for itself to the 'for the Other'" (i.e., from body for self to body for others)) "They are each the other side of the other." (p. 263). "The body sensed and the body sentient are the obverse and the reverse … as two segments of one sole *circular* course … which is but one sole movement in phases" (p. 138, my emphasis, see discussion of "Gestalt-circle," below).

Self other relationship then is mediated by an ongoing pre-attentive binding between body image and body schema, what Merleau-Ponty (1968) calls body for other(s) and body for self. Autoscopy may be a disruption of this binding that supports self-other relationship.

The Illusion of Self-Movement in OBEs: A Comparison of Theories

In contrast to Type I, a major feature of delusional, dreamlike Type II autoscopy, OBE's and FOP is the transformation of *one's own bodily experience*. This may involve a usurping (Type II), or reduction of mobility (FOP, as in sleep paralysis), or its increase (OBEs). In order to examine the overlap and difference with other theoretical approaches, I will focus on one problem, the experience of bodily movement during OBEs.

As clarification of the "hallucinatory" experience of increased mobility during OBE's, Metzinger (2005) gives the following example. While waiting for one's train to leave the station, one mistakenly experiences the train as moving. In actuality, it is the movement of the neighboring train viewed through the window that has caused the brief illusion. Metzinger explains that it is "a kinesthetic and proprioceptive hallucination, a non-veridical model of the weight and acceleration of your body, erroneously activated by your brain" (p. 61). The (illusory) proprioceptive input activated by the visual experience is meant to be suggestive of how the "hallucinatory" sensations of movement in OBEs come about. Metzinger's (2003, 2005) account relies on Gibson's (1958) thesis that vision provides the "kinaesthetic sense" in the example (in terms of the direction of movement-flow of the optic array). Metzinger (2003) writes, "The solution to this problem is to acknowledge that visual kinesthetic information, generally richer than mechanical kinesthetic information can overrule the second type (i.e., kinesthetic information) in cases of conflict because (citing Lishman and Lee 1973) 'vision is not only an exteroceptive sense, as is classically assumed, it is also an autonomous kinesthetic sense.'" (p. 491, my insertions).[37] Vision "affords" the kinaesthetic information independently of mechanical-muscular or proprioceptive kinaesthesis. Therefore, we take the cue that we are moving when the environment moves past us (as depicted by the sudden movement outside the train window). In order to resolve the conflict that the visual motion has introduced, vision for the moment "dominates" and an "accompanying kinesthetic-proprioceptive self-model" is activated

---

[37] Since 'mechanical kinesthesis' (along with vestibular, efferent and other sensory modalities) may be regarded as contributing to body schema, and body schema is primarily unconscious, it is not surprising that visual information would indeed seem "richer" to awareness than its nonconscious counterpart. However, it would be incorrect to conclude that the information (i.e., computations of body schema) are any less complex than body image (see the above discussion of the double who turns his head with the same (computational) complexity as the self in the case of polyopic heautoscopy).

leading to "a very brief hallucinatory episode" (Metzinger, p. 490). Therefore, the illusion of self-movement in OBEs arises primarily from the directionality of the perceptually given optic flow, which overrides a more informationally impoverished mechanical-kinesthetic system.[38] Moreover, "some persons undergoing an OBE truly are disembodied, thinking selves in a neurophenomenologically reduced version of the original Cartesian sense" (Metzinger 2003, p. 502). That is, the first person perspective – which may be minimal, paradoxically "unextended point in visual space" – is identified with visuo-attentional agency and provides the locus of phenomenal self during OBEs.[39]

Although valuable in its own right, the purpose of the current analysis is not to examine Metzinger's conceptually rich account of OBEs. It is rather to highlight some differences with the phenomenological approach. In describing the same train example, von Weizsäcker (1950), 1997)[40] argues that it is *not the priority of vision but rather its reduction in status that gives rise to the illusion.*

----

[38] The "illusory" experience of movement, for example, in virtual reality (and possibly, OBEs), however, does NOT solely depend on the direction of flow of the optic array. The perception of depth during "self-movement" in VR (and otherwise) also rests on motion parallax, the relative retinal image motion between objects at different distances (and therefore, egocentric computations). That is, closer objects should move more quickly in comparison to more distant objects relative to one's current movement speed. The objects' location relative to self during motion parallax is subserved by dorsal stream (area MT) in macaque monkeys and presumably, humans (Nadler et al. 2008).

[39] Metzinger states that the "consciously experienced egocentric frame of reference" – what I claim is neurobiologically implausible, see below – is the origin of the visual perspective and, during OBEs need not "occupy" any space at all for its mobile purview over the world and the body-object from which it has detached. In fact, some researchers (e.g., Bünning and Blanke 2005) share Metzinger's (2003, 2005) view that the vantage point of OBEs is an "egocentric visuo-spatial perspective." However, there are problems with this assertion: (1) Metzinger also writes, "the way in which OBE subjects move around in the currently active model of reality is not smooth, as in walking or flying, but occurs in discrete jumps from one salient point in the cognitive map to the next." That is, from Metzinger's own observation, it seems that the experience of movement in the OBEs is tracked in terms of an allocentric frame of reference or "cognitive map." Since the movement itself appears to be tracked in the allocentric coordinates, it is not clear how the subject would become aware of its egocentric perspective. We have repeatedly emphasized in the phenomenological analysis that the egocentric coordinates of the prospectively acting subject (e.g., the path structure in ocular motor response, or throwing/catching a ball) are already past and not directly available to the 'subsequent' reflexive awareness, "I move myself" (mediated by the 'faster' already occurring perception-action cycle, or Gestalt-circle between movement and perception). (2) Metzinger contends that "OBEs are like a perceptualized variant of reflexive self-consciousness" (2003, p. 502). However, this would require that the egocentric perspective be available to reflexive awareness, which (I have argued) is not supported by the phenomenological evidence.

[40] The philosophically minded sense-physiologist, neurologist and celebrated "founder" of psychosomatic medicine in Germany, Viktor von Weizsäcker, had tremendous impact on existential-phenomenological approaches in German-speaking psychiatry, as well as on Merleau-Ponty (described here) and Gadamer (1976, 1993). However, with the exception of a few researchers (e.g., Fuster 2006), he is practically unknown in English-speaking research (see Uhlhaas and Mishara 2007; Mishara 2004).

Rather than casting vision as "kinesthetic sense," von Weizsäcker (1948) proposes that *movement* (whether it be movement of self or objects) *is disruptive to perceptual organization*. When movement-speed exceeds a certain threshold, it disturbs perceptual coherence. In the train example, the externally generated movement is mistaken as self-movement because the train cabin, by extension, becomes one's realm of possible motoric performance or non-conscious body schema.[41] This may cause *quick adaptive movements* or adjustments of balance in one's own body when none are required (especially if one had been standing). [42] Here, the self and cabin are experienced as moving together (in the opposite direction to the other train leaving the station)[43] while *sacrificing* the more conscious visual scene information. Rather than visual (kinesthetic) information being the prime contributor to the illusion, it is the disturbance which movement introduces to the "balance" the subject maintains with its environment in the perceptual field. That is, *what is perceived as moving and what is perceived as stationary is relative to one's frame of reference.* In the train example, this is determined unconsciously in terms of the enactive body-schema (which struggles to maintain coherent relationship with the environment).

---

[41] Unlike body image, the (egocentric) computations of a non-conscious, short-lived, prospectively open body-schema exhibits an adaptive plasticity that is not confined to the self-enclosed (already past) unity of the predominantly visual body image. For example, my body schema momentarily becomes the baseball-glove catching the ball, the cane I use to walk, or the top-hat I am wearing as I automatically stoop while walking beneath a low bridge. The body schema is able to incorporate tools as if "our own effector (e.g., the hand) were elongated to the tip of the tool" (Maravita and Iriki 2004, p. 79).

[42] Similarly, von Weizsäcker and his assistants (reviewed by von Weizsäcker (1950)) observed during the experimental induction of vertigo that – in addition to the optokinetic nystagmus – rapid, adaptive movements with the head, torso and arms in the direction of the movement (often not perceived by the subject) play an organizing role in maintaining relationship with the disrupted perceptual experience. As with the train example, the illusion of self-movement (vection) during vertigo does not arise solely from the visual perception but from an embedded, enactive subject who maintains the coherence of his ongoing experience by making adjustments in the balance of perceptual and motor systems. See (Mishara in press a) for the impact of von Wiezsäcker's vertigo induction experiments on Conrad's and Binswanger's phenomenlogical approach to delusions in schizophrenia.

[43] For Metzinger (2005) expectation plays an important role in generating the illusory movement in the train example: "At the same time there was a state of general physical and emotional arousal, accompanied by an unconscious state of expectancy about what is very likely going to happen next, and very soon." (p. 61). If expectation plays such a role in 'selecting the interpretation' in a "system that always tries to maximize overall coherence," how is it that there is no incongruency felt by the fact that the illusory self-movement of one's own train (i.e., going backwards) is precisely opposite to the direction expected? It is precisely because the movement is unexpected that it presents a momentary "crisis" for the subject (von Weizsäcker 1950, see also below). It therefore elicits a compensatory involvement of the motoric body schema reflected (in part) by the unnecessary movements of the subject (especially if standing in the train cabin) to maintain balance. Critically, Metizinger's use of the term "coherence" depends on a different concept of subjectivity (as self-model) than von Weizsäcker's (see *Dissociating Mind and Subjectivity*, below).

Following this line of thought, the various forms of autoscopy may be variants of compromise between perception and movement with the purpose of maintaining coherence in an otherwise threatened self-world relationship. Various *components of self are preserved or even exaggerated* while sacrificing others. OBE experiences may be the uncoupling of body schema (as I) from body image (as me, now other), whereby: (1) the mobility of the subject is preserved by consigning (restricting) the immobility or "paralysis" (e.g., Case 4) to the body image (now detached); (2) space is re-organized relative to the unencumbered, mobile body schema.

Perception-Action Cycle and Self-Other Relationship

When the integrity of the perceptual Gestalt is disrupted due to excessive movement, von Weizsäcker (1950) proposes that we effortlessly and spontaneously apply "strategies" that preserve the coherence of the perceptual field as a totality (as much as this is possible). It is a totality that includes the embedded subject in its ambiguous relationship to this field as both perceiver and agent. Therefore, *the (perceived) continued existence of the subject depends, in part, on the preserved coherence of the perceptual field*. This is maintained by sacrificing some features while preserving others. For example, the whir of a propeller blade occurs at speeds that we are no longer able to perceive the individual blades. In this case, movement disrupts the integrity of the perceptual Gestalt. The disturbance, however, is limited by the fact that the individual blades appear to fuse. Critically, we learn not take these perceptual changes "seriously." The spontaneous emergence of a new organization assigns functional significance of each part to every other part by setting the conditions for each locus in maintaining inner coherence of the system (Gurswitsch 1966). The strategies to maintain coherence by reorganizing preserved and sacrificed aspects are not consciously generated but appear to emerge effortlessly and automatically in the service of preserving the self-world relationship as meaningful. The resulting transitions in perceptual organization are rapid and ballistic in that each new organization appears to emerge on its own without precedent in the prior organization (what von Weizsäcker (1950) calls "improvisation"). *Delusions, dreams and autoscopy are ways of preserving this coherence at different degrees of disruption to the perceptual Gestalt, mobility and/or consciousness* (see Uhlhaas and Mishara 2007). Conrad, Binswanger, Blankenburg and other phenomenological psychiatrist held the view that delusions in schizophrenia resemble dreaming in that the "objects" expreienced in both dreaming and delusions are based on incomplete perceptions or meanings - what Conrad calls the Pre-gestalt (Vorgestalt) - which nevertheless fascinate the subject and from which the subject is unable to detach, very much like the hallucinatory type II, dream-like autoscopy (see Mishara in press a).

In summary, the relationship between the subject and experience is *a fragile balance between conscious perception and non-conscious movement*. The formation of the (perceptual) Gestalt (whereby vital contact with the environment is maintained) occurs in a "circular" movement (between the mutually exclusive and

yet, inseparable moments of movement and perception).[44] The ongoing "compromise" between perception and movement (in a single "act" of preattentive binding) only becomes accessible to experiment by disrupting it.[45]

The hallucinatory sense of movement in OBEs, the double's echopraxis in Type II autoscopy, or the doubling of self during oppressive experiences of FOP during sleep paralysis, may be various ways of preserving coherence or balance between the subject and the environment.

Dissociating Mind and Subjectivity

As previously indicated, it is currently fashionable to propose degrees of disembodiment as a cognitive-neural mechanism underlying the different forms of autoscopy. In a commentary to Metzinger (2005b), however, Gallagher (2005) writes, "there is no such thing as real disembodied experience. Full-blown pre-reflective embodiment should not be thought of as simply a body image that could be generated by neuronal stimulation, or a simulated functional body-schematic system that could be activated in an exclusively neural matrix. The lived body, the body I live, is the real biological body, and if it were taken away, the life-support system that would have to replace it would necessarily be a real complex system that I could live and experience in the same way as I live and experience my body" (p. 7).

Metzinger (2005b) counters that "Gallagher in describing the lived body as 'the body I live' introduces *a distinction between himself and his body*, the relationship between the two being that he himself, Shaun Gallagher, 'lives it.'… The self-model theory is free of these underlying *Cartesian* intuitions... when you refer to yourself *saying I* you refer to the system as a whole, including your brain, body, self-model, history and social context-but you do so in a very special displaced manner: by using the content of your PSM (phenomenal self model) as an intermediary in the act of self-reference most of the time without noticing this fact" (p. 6, my parenthetical insertion and emphases).

---

[44] The Gestalt-circle occurs according to a "revolving door principle" (*Drehtuerprinzip*): "Each act is perception and movement. However, I am unable to perceive in my perception the movement that made it possible. Conversely, I am unable to access in the movement the perception that guides it. … Movement and perception stand in a relationship of mutual concealment" von Weizsäcker (1950), p. 200, my translation). The relationship is circular in the sense that one is "unable to ever establish where the relationship begins or ends" (von Weizsäcker, 1997 (1933), p. 26; my translation).

[45] "In my view, the method most suitable for the matter described here never makes the Gestalt itself available for analysis but always the limits of its appearing and disappearing, that is, the conditions of its formal principles and not its content" (von Weizsäcker 1948, p. 11). While phenomenological purists may object that this method is "non-phenomenological," it is, in actuality, very close to Husserl's phenomenological method. It examines meaningful coherence of the Gestalt in terms of its *limits* (i.e., when the Gestalt loses its form or is no longer meaningful (e.g., vertigo-induction experiments von Weizsäcker 1950) and thus, resembles Huserl's second step of eidetic imaginative variation outlined above.

The existential-phenomenological approach to self and its disruption in autoscopy (presented here) differs nontrivially from Metzinger's (2006b) response on several points: (a) it is not possible to reduce self to self-reference, a self-model or a self-representation; (b) the human relationship to body and the resulting vulnerability to disorders of self-experience (e.g., autoscopy) *is fundamentally ambiguous and ambivalent* (i.e., what Metzinger refers to as "Cartesian intuitions"; (c) the "body I live" (living subjectivity) is something much more fundamental than "saying I."

a. As already indicated, cognitive neuroscientists (e.g., Gusnard 2005) and philosophers often confuse self-awareness in self-referential processing or representational content about self, i.e., having a self (a self-enclosed entity, e.g., Metzinger's phenomenal self model), with being a self, prospectively open to its own future. Any reference to (or representation of) self as entity is *pars pro toto* (taking a part for the whole) and is after the fact. In existentialist terms, the human self is "condemned" to self-transcendence, i.e., condemned to prospectively transcend each of its current representations of self (Mishara 2007a).

b. Human cognitive development is characterized by a fundamental "common sense dualism" between mind and body (Bloom 2004; Bloom and Weisberg 2007). Despite well-meaning efforts of researchers and philosophers to view the mind and brain as ultimately the same, this tendency is nearly intractable. In my view, this dualism is reflected in the opposition of body image (body as object) and body schema (body as subject) (Mishara 2005). The fact that we are able to take *both* an internal-vital (i.e., proprioceptive-interoceptive) and external (exteroceptive, social-objectifying) relationship to our own bodies is the precondition for any vulnerability to the disruption of self-experience in neuropsychiatric disorders and anomalous conscious states (Mishara 2005). The German phenomenologist, Plessner (1975) uses the phrase *Leib im Koerper* (i.e., body as subject (*Leib*) as it is lodged in but not coincident with body as object, *Koerper*). That is, my embodied being-in-the-world as self-transcendent is ec-static, prospectively *open and vulnerable* to the not-yet-known in a way that extends beyond my experience of having a self-enclosed body image. (For review of Plessner's distinction, see Mishara in press c).

c. Closely related to the previous points, the first person "spoken I" (which Metzinger refers to) should not be confused with what is meant here by "living subjectivity" (Gallagher's "body I"). The first person spoken "I" is formal and empty.[46] With regard to the body subject, von Weizsäcker (1948) writes: "In crisis, what we experience as mind (*das erlebbare Psychische*) confronts its own limits. Vertigo, weakness, loss of consciousness and coma are capable – in their extreme forms – of dissolving what we call mind (as coherent organization) into

---

[46] Gallagher (2004) writes: "… self-knowledge conceived as this first person access is immune to error through misidentification … when I say "I think X," I can be mistaken about X, but I cannot be mistaken about to whom the 'I' refers. Self-reference is guaranteed. For this, however, it is all the more impoverished. It remains a formal principle, nothing more than a transcendental index that accompanies every experience in life that is meaningful (Kant)." (p. 8).

chaos. When a living being thereby loses this inner appearance (of mind), it does not mean that this being has lost its life, individuality, material basis or even its form. Just as Schopenhauer, E. v. Hartmann and Freud had made a decisive turn by challenging the equation of consciousness with mental, so we *must also give up the equation of mental with subjective.* This means that a person in coma, who is deprived of consciousness, and an organism, which may not have a specific experience of mind, nevertheless *comport themselves to their environment as subjects.* Moreover, this relationship is *not* representable as being either predominantly mental or physical (physiological). As a merely negative determination, this '*not*' would be uninteresting. When in crisis, however, the subject is not only another name for this '*not*' it is also precisely what is threatened or preserved in the unity of the organism. We only first really notice our own subjectivity when it is threatened to dissolve in crisis. … The subject is not a firm possession but must be acquired anew at each moment to 'possess' it. … (T)he unity of the subject is only first constituted in its ongoing incessant reestablishing itself in crisis and its own infirmity." (1950, pp. 172–173, my translation and emphases). The phenomenological-psychiatrist Binswanger identifies von Weizsäcker's concept of the *subject* as *the hidden unity* of the "Gestalt-circle" between movement and perception (and by extension, self and other) with the existential concept of being-in-the-world as an ongoing process of (inevitable) self-transcendence. *Being* (rather than *having*) a self is the ongoing vulnerability of being oriented towards the future, the not-yet-known.

Metaphor as Symbolic Self-transcendence in Autoscopy

Case 8

After the distressing break-up of her marriage, a 47-year-old woman, reports that in a moment of despair while at her job in a large Hotel-kitchen, she entertains emerging thoughts of suicide, which she had otherwise successfully suppressed. Terrified, she goes outside and sits on a bench. At this moment, she sees that a woman has hanged herself on a neighboring tree. Startled, she realizes that the woman is herself wearing her usual street clothes. She stares at the apparition for approximately one minute. After shutting and reopening her eyes, the figure vanishes (translated and paraphrased from Arenz 2001).

Although suffering from depersonalization, depression and nightmares following the break-up, the patient has no previous history of personality disorder, neurologic or psychiatric disease. This case is an elegant indication that the autoscopic double may take on symbolic significance for the subject (without effort or awareness as if in a dream).[47] Arenz (2001) interprets its "symbolic" character "as expression of the patient's inner fear of death elicited by her own suicidal thoughts projected onto the

---

[47] I am indebted to Peter Brugger for bringing my attention to this case as an example of symbolic (he)autoscopy.

outside world" (p. 378, my translation). That is, from a psychoanalytic viewpoint, projection functioned as a defense that helped the patient cope with the otherwise intolerable anxiety elicited by her own suicidal thoughts.

The phenomenological approach developed here proposes an alternative to the psychoanalytic interpretation. When controlled processing relaxes, metaphoric images of self spontaneously arise (e.g., hypnagogic images, spontaneous metaphors in narratives) (Silberer 1909; Mishara 1995). These reflect the self in its dual movement in time as advancing towards the future and letting go of the past, what Husserl calls "I move myself" as the core of self-transcendence in time (Mishara 2007a).[48]

Binswanger (1965, 1957) observes that self-transcendence (the "I move myself") is the condition for the ability to distance from one's current experience. Moreover, it is compromised in acute psychosis as well as during dreaming and other anomalous conscious states. As a result, patients with schizophrenia sometimes delusionally refer to themselves in the most inhuman, "thinglike" terms, e.g., as a "machine," "computer," or "apparatus" whose sole function is to "register" impressions. This concretization of metaphors of self is nevertheless an implicit way of preserving a (minimal) self in its compromised ability to transcend the present perspective (Mishara 2007a). That is, the metaphoric description of self as a "registering apparatus" at once testifies to the patient's compromised ability to transcend current experience but, at the same time, preserves a distance (no matter how minimally) to this experiencing by enabling the patient to (metaphorically) describe, and thus, transcend the experience by use of metaphor (no matter how concretely interpreted).

A similar process occurs in autoscopy. In Case 8, the autoscopic image anticipates and thereby usurps the patient's frightening intention to commit suicide. Here, the prospective self is preserved by letting go of the *me* that the *I* just was (the momentary "I" who entertained suicide). The agency of body schema is preserved either by having the double assume it (Type II autoscopy) or by having the subject detach from a rather inert and 'inanimate' body image (OBEs). In each of these cases, the self is preserved by a reorganization of the experience.

Do the Neuroanatomical Correlates of Autoscopy Support the Phenomenological Theory?

As the other approaches, phenomenology proposes its own hypotheses (Mishara 2007b) about the underlying cognitive and neural mechanisms of autoscopy. I will

---

[48] We experience our consciousness as an obligatory displacing itself with each new now. Nevertheless, we do not have reflective access to this process which is fundamentally self (Mishara 2007b). As Husserl writes, "the streaming is always ahead (*im voraus*) …" Any representation of self is already past (*having a self*), with its own *closure* (i.e., encoded in allocentric coordinates). This *retrospective* unity (body-image) cannot replace my current openness to a future, an incomplete openness with each new now moment. Self as this process of self-transcendence is reflected in Sartre's (1966) famous phrase, "existence precedes essence" where essence (citing Hegel) is "what is already past": "Wesen ist, was gewesen ist" (See Mishara 2004, 2007a).

briefly review findings regarding the underlying neural substrates of autoscopy to examine whether they lend support to the phenomenological hypotheses.

In Type I autoscopy, the subject retains first person awareness and encounters a doubled body image, a second me. Consistent with the view that Type I autoscopy could involve an uncoupling of body image and body schema in which the me but not the I is impacted, it is predominantly a unimodal, visual experience and echopraxia (when present) is symmetric in allocentric coordinates as if in a mirror. While sometimes occurring in healthy individuals, it is most frequently associated with lesions predominantly to the occipital-parietal areas with the double appearing on the side of hemiopnia (Blanke and Mohr 2005). While such lesions could obstruct either or both dorsal and ventral visual processing streams, the abnormality appears to be largely confined to body image (thus implicating predominantly the ventral stream).

There are several reasons for thinking that Type II autoscopy and OBE's involve primarily a disturbance to body schema. Unlike Type I autoscopy, the experience is multimodal. Vestibular-hallucinations and somatosensory/proprioceptive disturbances are comparatively frequent. When present in Type II autoscopia, echopraxia is implemented with the contralateral limb or half of the body (i.e., in egocentric coordinates) and may even anticipate the thoughts or actions of the subject. *The egocentric coordinates of an on-line body schema are computed from the integration of multiple sources, including afferent information from sensory input, proprioceptors and the vestibular system.* Interestingly, the tempoparietal junction (TPJ) as well as surrounding superior-temporal gyrus and tempoparietal cortex have been implicated in both Type II autoscopy and OBE's. The TPJ is thought to be involved in the integration of visual, somatosensory/proprioceptive and vestibular information. Although still controversial, the TPJ and proximal posterior insula are thought to be the homologue in the human brain of the "vestibular cortex" reliably identified in the macaque and marmoset brains to receive the majority of cortical projections from the vestibular system (Brandt and Dieterich 1999; Grüsser et al. 1990). (For location of TPJ, see Fig. 2).

There is further reason to believe that the TPJ is implicated in disturbance to body schema and/or the coordinated interplay of body image and body schema found in Type II and OBE's. In Type II autoscopy, which I have claimed to be dream-like, the double assumes the active role. As the hand being touched (in Merleau-Ponty's "double-sensation" experiment), the embodied subject loses its active grip on the world. To the extent that the double mimics or even anticipates the subject's own actions and wishes, the double usurps the body schema, and the conscious self becomes identified with the conscious body image (as object from the double's point of view).

In OBE's, the body schema is preserved (in a so-called first-person point of view) but is experienced as separate from the body image (which is often seen as inanimate from the self's current hovering perspective). As previously noted, this experience is better described as a mode of human embodiment rather than as disembodied.

In a PET study of a reciprocal imitation paradigm (Decety et al. 2002), the *left* inferior parietal lobule (IPL, i.e., angular gyrus (BA 39) was activated when subjects imitated the other, while the *right* homologous region was associated with being imitated by the other. A proximal area in the posterior part of the superior temporal
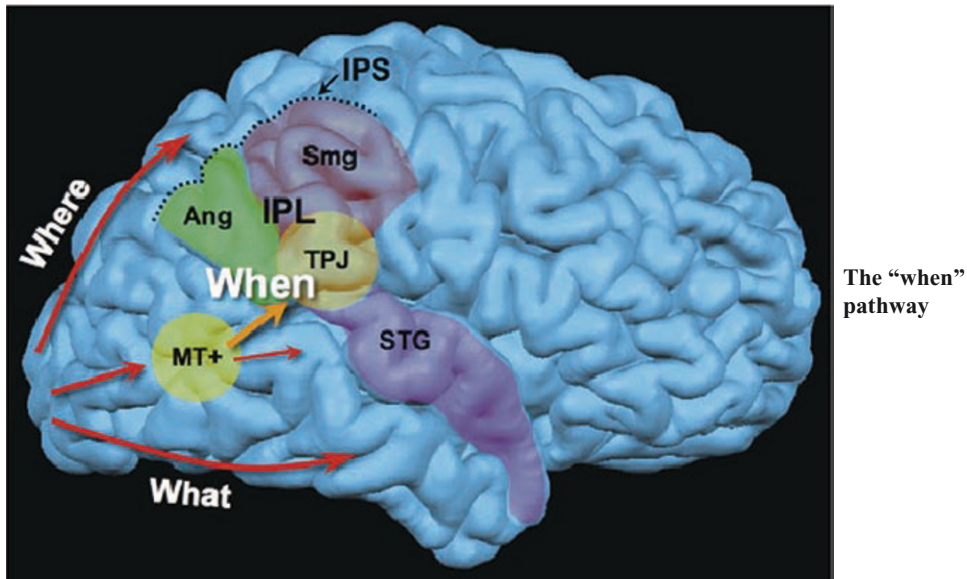
The "when" pathway

Fig. 2 The *when* pathway. This pathway is lateralized in the right hemisphere. Information from the primary visual cortex (V1) travels along the dorsal pathway (spatial perception, determining *where* objects are) or the ventral pathway (object recognition, determining *what* objects are), according to the classical subdivision that has been proposed based on animal models. A third pathway coming from V1 is dedicated to using time information to identify objects (e.g. determining when objects appeared or disappeared). Here, the temporoparietal junction (TPJ; considered the most common substrate of neglect) is identified as a core anatomical locus, within the inferior parietal lobe (IPL); however, the *when* pathway is likely to include a bigger network of areas, including the right angular gyrus (Ang), the supramarginal gyrus (Smg) and the posterior superior temporal sulcus (included in the superior temporal gyrus, STG). All these areas are often involved in the cortical lesion of right parietal patients. The intraparietal sulcus (IPS) separates the IPL from the superior parietal lobe (not labeled). The middle temporal area MT+ is reported in yellow (also called the motion area, highly specialized in detecting and discriminating moving stimuli) (Reprinted from *Trends in Cognitive Sciences,* 11, Battelli, L., Pascual Leone, A., and Cavanagh, P. The 'when' pathway of the right parietal lobe, Copyright (2007), 204–210, with permission from Elsevier)

gyrus (STG) was also found to be active when subtracting the condition of being imitated from the condition of imitating the other. The latter area is implicated in biological motion perception and receives input from both the ventral and dorsal visual streams as well as the vestibular system. These findings are pertinent when one considers that Type II autoscopy is more associated with abnormalities of the *right TPJ* and surrounding areas, whereas OBEs are more associated with abnormalities or stimulation of the *left TPJ* and surrounding areas, including angular gyrus (BA 39) and posterior STG (e.g., Lopez et al. 2008).

In other words, *both Type II autoscopy and OBE may the expression of the uncoupling of body schema and body image, but in opposite directions and involving different nodes in a self-other network.* In analogy to Decety et al.'s (2002) findings, one may say that the Type II double is the *leader* and the subject is the follower during reciprocal imitation, i.e., that the *double takes over the I and*

becomes the (visible) embodiment of the body schema. Therefore, both passively imitating the other in the reciprocal imitation paradigm and Type II autoscopy, which may involve autoscopic echopraxia, are associated with activity in *right TPJ*. It is quite different in OBE's. Here, the subject, the I, is relatively freed from the encumbrances of *having a body*, or body image, which is often seen from above. In sleep paralysis, OBE's are associated with a freeing or pleasant sensation as opposed to the oppressive FOP (which are rarely in the same episode). Here the I (or body schema) takes on the active role and is therefore associated with the *left TPJ* (i.e., the active "leader" role in the Decety et al. (2002) experiment).

As further support of right TPJ and proximal areas being implicated in Type II autoscopy, failure to attenuate activation of the inferior parietal lobe (IPL), especially the *right IPL* (neighboring the TPJ, see Fig. 2), leads to *the attribution of movements of one's own body to external agency* both during hypnosis in healthy subjects (Blakemore et al. 2003) and in patients with schizophrenia who experience delusions of alien control (Spence et al. 1997). The same area becomes activated when subjects attribute actions to someone else as opposed to themselves. The right IPL has been proposed to be critical for taking the other person's perspective or when subjects mentally simulate actions from the other's perspective (Vogeley et al. 2004; Ruby and Decety 2001; Vogeley et al. 2001). Thus, failure to attenuate right IPL (in delusions of control) or proximal TPJ may underlie the illusory attribution of one's own actions (or body schema) to an external agent in both delusions of control and the experience of a *Doppelgänger* usurping the "real self" in Type II autoscopy.

In FOP, the "double" is felt but not directly seen. It may felt to be related to the self (as in Brugger et al. 2006) or as foreign, even hostile to the self (as in sleep paralysis).[49] In FOP, the body schema is disrupted by being shared with another. This may be due, for example, to the experienced loss of movement in sleep paralysis (which is recovered in part by the illusory mobility of OBEs). Moreover, without seeing the presence, the subject somehow 'knows' that the shadowy presence imitates the subject's own movements. By stimulating the *left TPJ* through subdural electrical stimulation, Arzy et al. (2006) were able to induce the feeling of a shadowy presence that imitated the subject's movements. When lesions are present, feeling-of-a-presence is often associated with parietal cortex (i.e., proximal to TPJ, see Fig. 2). These findings suggest that FOP is also a variant of disruption of body schema. However, unlike Type II autoscopy, the subject still maintains the sense of being the primary locus for the self but, unlike OBEs, the double alien-self is not experienced as a body object but as another subject.

That is, in FOP and Type II autoscopy, one's agentic self (body schema) is compromised but also preserved by being shared with, or even attributed to the double,

---

[49] With regard to intruder (FOP) and incubus hallucinations, Cheyne (2003) remarks, "there also appears to be a link between visual association areas, object identification (implicated by the ventral stream involvement), and limbic activity." (p. 165). This concurs with our hypothesis that autoscopy-related disorders involve a disruption of relationship between body image and body schema, mediated by ventral and dorsal visual processing streams, respectively (see also Mishara 2005, 2007a).

respectively.[50] Because the pre-attentive binding of the subcomponents of self relies on precise temporal relationships between the neural pathways subserving them, the disruption in a (putative) "when" pathway (converging on TPJ, see Fig. 2), which also depicts the ventral and dorsal visual processing streams, i.e., the "what" and "where" pathways; see, e.g., Galati et al. 2001, for putative neural correlates in humans for processing body-centered, egocentric coordinates) may be a possible factor leading to autoscopic experiences.

Conclusions

Given that this is a general overview, it is only possible to suggest some areas of overlap and difference between cognitive and phenomenological theories and future research directions. By taking some views expressed by the cognitive science/neuroscience researchers, Chris Frith and colleagues, I will point out some areas of overlap as well as difference.

(a) Frith (1995) remarks: "the major mistake of most theories of consciousness is to try to develop an explanation in terms of an isolated organism." This claim resonates with the phenomenological position developed here that self-other relationship (i.e., "turn taking") is a revolving door of mutual exclusion provided by the human subject as a kind of retrospective/anticipatory "structure" (mediated by a perception action cycle) which, in turn, is applied to one's relationship to oneself. Embodied subjectivity is embedded in intersubjectivity.

(b) For Frith (1995), the non-conscious motor system (what I have been describing as on-line body schema) enables the pre-attentive binding "between vision and movement permitting (e.g.) grasping … The motor representation provides an absolute egocentric calibration of egocentric space … I propose, as a general principle, that consciousness contains only those representations that are coded independently of egocentric coordinates. *Thus the model of the world that is represented in our consciousness is, as far as this is possible, independent of our point of view"* (p. 682, my emphases). We have repeatedly emphasized in the phenomenological analysis that the egocentric coordinates of the prospectively acting subject are already past and not directly available to reflective awareness and, as far as we can know, pre-reflective awareness as well (see Mishara in press a,d for the reflective and experimental conundrums). Patients with frontal lobe lesions are sometimes unable, for example, to stop putting on several pairs of eye-glasses, one on top of the other ("utilization behavior," Lehrmitte, 1983). This indicates that the egocentric or body-centered reference frame as shaping a space of "affordances" calling for many more possible

---

[50] See, for example, the description of the patient (Case 4) who experiences his own body as "paralyzed," but whose vantage point alternates between the double and immobile body self.

actions than we can be aware at any given moment is, under "normal" circumstances, inhibited and unconscious. "I cannot see myself in movement, witness my own movement" (Merleau-Ponty, 1968, p. 254), that is, except in terms of its *effects* in an embodied space which has been already transformed *by a just past perception action cycle* (or more appropriately, *action perception cycle*) into a consciously experienced allocentric frame of reference, i.e., "independent of our own point of view." This is a subtle point often overlooked by philosophers or psychologists not acquainted with the neuroscience, but critical to the current phenomenological analysis of the different forms of autoscopy.

(c) During observation of another's actions, social interaction and imitation, the motor system is implicated (Wolpert et al. 2003). In cooperative efforts or reciprocal imitation, self and other anticipate (i.e., predict) each other's movements in terms of complementary inverse and forward models of motor control. Imitating another's movements or actions is seen as just such a problem of converting the perceptual experience of another's action into motor commands necessary to perform the action (inverse model). The effects of this action are then again converted to its predicted sensory consequences in terms of its efference copy (forward model). Presumably, such mutual anticipation by reversal of the other's agentic goals would not be conscious and allow for both cooperative and competitive actions. Commenting on such a theory, Miall (2003) writes: "This two-way process could also allow an observer to track another's hand actions with predictive eye movements, the gaze shifts anticipating the other's hand motion with the same advance as seen when tracking one's own actions. It could allow cooperative actions such as shaking hands or dancing, or it could allow successful competitive actions, such as when we both attempt to grab the same bit of apple."[51] Proponents of an action-understanding theory of (obligatory) covert imitation have proposed that exactly this kind of motor control forward model is co-opted for generating the predicted sensory consequences of others' actions as if they were one's own (e.g., Blakemore and Decety 2001).

Herein lies the difference with the above position: In the phenomenological theory, the reciprocal imitation (as an ongoing (automatic and obligatory) exchange of body schema and body image between self and other in a 'precisely calibrated dance') is not merely occasioned at certain moments of cooperation or competition with others. It is rather the very structure of the self (i.e., to be simultaneously other but in a way hidden to oneself) and it is this structure of self (as self-other relationship), which is ultimately vulnerable to its own disruption in the experience of

---

[51] Similarly Gadamer (1976) (citing von Weizsäcker (1950)), writes about playing an athletic game, which, by definition, is both cooperative and competitive: "the game is … the formation of the movement as such, which in an unconscious teleology subordinates the attitude of the individuals to itself… neither partner alone constitutes the real determining factor, rather, it is the unified movement as a whole that unifies the activity of both. We can formulate this as a theoretical generalization by saying that the individual self, including his activity and understanding of himself, is taken up into a higher determination that is really the decisive factor" (pp. 53–54).

doubles (Mishara in press d). The similarities and differences between cognitive science and the phenomenological theory should be further studied with regard to their application to the disruption of self-other relationship in neuropsychiatric disorders and anomalous conscious states.

References

Ach N (1905) Über die Willenstätigkeit und das Denken. Vandenhoeck & Rubprecht, Göttingen

Arenz D (2001) Heautoskopie. Doppelgängerphänomen und seltene Halluzination der eigenen Gestalt. Nervenarzt 72:376–379

Arzy S, Seeck M, Ortigue S, Spinelli L, Blanke O (2006) Induction of an illusory shadow person. Nature 443:287

Battelli L, Pascual-Leone A, Cavanagh P (2007) The 'when' pathway of the right parietal lobe. Trends Cogn Sci 11:204–210

Binswanger L (1957) Schizophrenie. Neske, Pfullingen

Binswanger L (1965) Wahn. Neske, Pfullingen

Blakemore S-J, Decety J (2001) From the perception of action to the understanding of intention. Nat Rev Neurosci 2:561–567

Blakemore S-J, Oakley DA, Frith CD (2003) Delusions of alien control in the normal brain. Neuropsychologia 41:1058–1067

Blanke O, Mohr C (2005) Out-of-body experience, heautoscopy, and autoscopic hallucination of neurological origin. Implications for neurocognitive mechanisms of corporeal awareness and self-consciousness. Brain Res Brain Res Rev 50:184–199

Blanke O, Landis T, Spinelli L, Seeck M (2004) Out-of-body experience and autoscopy of neurological origin. Brain 127:243–258

Blankenburg W, Mishara AL (trans) (2001) First steps toward a 'psychopathology of common sense'. Philos Psychiatr Psychol 8:303–315 Originally published in 1969

Bloom P (2004) Descartes' baby. Basic Books, New York

Bloom P, Weisberg DS (2007) Childwood origins of adult resistance to science. Science 316(5827): 996–997

Brandt T, Dieterich M (1999) The vestibular cortex: its locations, functions, and disorders. Ann N Y Acad Sci 871:293–312

Bruce R (1999) Astral dynamics. A new approach to out-of-body experience. Hampton Roads, Charlottesville, VA

Brugger P (2002) Reflective mirrors: perspective-taking in autoscopic phenomena. Cogn Neuropsychiatry 7:179–194

Brugger P (2007) Hostile interactions between body and self. Dialogues in Clinical Neuroscience 9:210–213

Brugger P, Agosti R, Regard M, Wieser H-G, Landis T (1994) Heautoscopy, epilepsy, and suicide. J Neurol Neurosurg Psychiatry 57:838–839

Brugger P, Regard M, Landis T (1996) Unilaterally felt "presences": the neuropsychiatry of one's invisible doppelganger. Neuropsychiatry Neuropsychol Behav Neurol 9:114–122

Brugger P, Regard M, Landis T (1997) Illusory reduplication of one's own body: phenomenology and classification of autoscopic phenomena. Cogn Neuropsychiatry 2:19–38

Brugger P, Regard M, Landis T, Oelz O (1999) Hallucinatory experiences in extreme-altitude climbers. Neuropsychiatry Neuropsychol Behav Neurol 12:67–71

Brugger P, Blanke O, Regard M, Bradford DT, Landis T (2006) Polyopic heautoscopy: case report and review of the literature. Cortex 42:666–674

Buckner RL, Carroll DC (2007) Self-projection and the brain. Trends Cogn Sci 11:49–57

Bünning S, Blanke O (2005) The out-of body experience: precipitating factors and neural correlates. Prog Brain Res 150:331–350

Cheyne JA (2003) Sleep paralysis and the structure of waking-nightmare hallucinations. Dreaming 13:163–179

Cheyne JA (2005) Sleep paralysis episode frequency and number, types, and structure of associated hallucinations. J Sleep Res 14:319–324

Christodoulou GN (1978) The syndrome of subjective doubles. Am J Psychiatry 135:249–251

Claesges U (1964) Edmund Husserls Theorie der Raumkonstitution. Phaenomenologica. Martinus Nijhoff, Den Haag

Conrad K (1933) Das Koerperschema. Eine kritische Studie und der Versuch einer Revision. Zeitschrift fiir die gesamte Neurologie und Psychiatrie 147:346–369

Conrad K (1953) Über ein eigenartiges Spiegelphantom. Heautoskopisches Phänomen als Dauerzustand bei Hypophysentumor. Nervenarzt 24:265–270

Conrad K (1958) Die beginnende Schizophrenie. Stuttgart, Thieme Verlag

Conrad K (1960) Die symptomatischen Psychosen. In: *Psychiatrie der Gegenwart*, Vol. II, Ed. by H. W. Gruhle, R. Jung, W. Mayer-Gross and W. Mueller, Berlin, Springer, pp. 369–436

Damas Mora JMR, Jenner FA, Eacott SE (1980) On heautoscopy or the phenomenon of the double: case presentation and review of the literature. Br J Med Psychol 53:75–83

De Preester H (2007) The deep bodily origins of the subjective perspective: models and their problems. Conscious Cogn 16(3):604–618

Decety J, Chaminade T, Grèzes J, Meltzoff AN (2002) A pet exploration of the neural mechanisms involved in reciprocal imitation. Neuroimage 15:265–272

Denning TR, Berrios GE (1994) Autoscopic phenomena. Br J Psychiatry 165:808–817

Devinsky O, Feldmann E, Burrowes K, Bromfield E (1989) Autoscopic phenomena with seizures. Arch Neurol 46:1080–1088

Ehrsson HH (2007) The experimental induction of out-of-body experiences. Science 317:1048

Eichenbaum H, Yonelinas AP, Ranganath C (2007) The medial temporal lobe and recognition memory. Annu Rev Neurosci 30:123–152

Engerth G, Hoff H (1929) Ein Fall von Halluzinationen im hemianoptischen Gesichtsfeld. Beitrag zur Genese der optischen Halluzinationen. Monatsschrift für Psychiatrie und Neurologie 74:246–256

Feinberg TE (2000) Altered Egos: How the Brain Creates the Self. New York, Oxford University Press

Féré C (1891) Note sur les hallucinations autoscopiques ou spéculaires et sur les hallucinations altruistes. C R Séanc Soc Biol 3:451–453

Foa EB, Riggs DS (1993) Posttraumatic stress disorder and rape. In: Oldham JM, Riba MB, Tasman A (eds) Review of psychiatry, vol 12. American Psychiatric, Washington, DC, pp 273–303

Frith CD (1995) Consciousness is for other people. Behav Brain Sci 18:682–683

Fuchs T (1995) Coenaesthesie. Zur Geschichte des Gemeingefühls. Z Klin Psychol Psychopathol Psychother 43:103–112

Fuchs T (2006) Ethical issues in neuroscience. Curr Opin Psychiatry 19:600–607

Fuster JM (2006) The cognit: a network model of cortical representation. Int J Psychophysiol 60:125–132

Gadamer HG (1976) In: Philosophical hermeneutics (trans: Linge DE). University of California Press, Berkeley

Gadamer HG (1993) Truth and method, 2nd revised edition (trans: Weinsheimer J, Masrshall D). Continuum, New York. (Original work published 1960)

Gallagher S (2004) Ways of knowing the self and the other. In Gallagher S, Watson S (eds) Ipseity and alterity: interdisciplinary approaches to intersubjectivity. Publications de l'Université de Rouen, Rouen, pp 7–29

Gallagher S (2005) Metzinger's matrix: living the virtual life with a real body. Psyche 11:9

Gallagher S, Cole J (1995) Body schema and body image in a deafferented subject. J Mind Behav 16:369–390

Gibson JJ (1958) Visually controlled locomotion and visual orientation in animals. Br J Psychol 1958:182–194

Goethe JW (1790) Versuch die Metamorphose der Pflanzen zu erklären. Carl Wilhelm Ettinger, Gotha; 1984 facsimile. Acta humaniora der Verlag Chemie GmbH, Weinheim

Green CE (1968) Out-of-body experiences. Institute of Psychophysical Research, Oxford

Grüsser OJ, Landis T (1991) The splitting of "i" and "me": heautoscopy and related phenomena. In: Grüsser OJ, Landis T (eds) Visual agnosias and other disturbances of visual perception and cognition, vol 12. Macmillan, London, pp 297–303

Grüsser OJ, Pause M, Schreiter U (1990) Localization and responses of neurones in the parieto-insular vestibular cortex of awake monkeys (Macaca fascicularis). J Physiol 430:537–557

Gurwitsch A (1966) Studies in phenomenology and psychology. Northwestern University Press, Evanston, IL

Gusnard DA (2005) Being a self: considerations from functional imaging. Conscious Cogn 14:679–697

Haggard P, Aschersleben G, Gehrke J, Prinz, W (2002) Action, binding and awareness. In B. Hommel, & W. Prinz (Eds.), *Attention and Performance XIX: Common Mechanisms in Perception and Action*, pp. 266–285. Oxford, England: Oxford University Press

Head H (1920) Studies in Neurology, 2. Oxford University Press, London

Head H, Holmes G (1911) Sensory disturbances from cerebral lesions. Brain 34:102–254

Hécaen H, Ajuriaguerra J (1952) L'Heautoscopie, Meconnassiances et Hallucinations L'Héautoscopie. Corporelles, Masson, Paris

Heidegger M (1927) Martin Heidegger. Sein und Zeit. Niemeyer, Tubingen

Held K (2003) Husserl's phenomenology of the life-world. In: Donn W (ed) The new Husserl: a critical reader. Indiana University Press, Bloomington, pp 32–62

Hoffman RE (2007) A social deafferentation hypothesis for induction of active schizophrenia. Schizophr Bull 33:1066–1070

Hoffman RE, Hawkins KA, Gueorguieva R, Boutros NN, Rachid F, Carroll K, Krystal JH (2003) Transcranial magnetic stimulation of left temporoparietal cortex and medication-resistant auditory hallucinations. Arch Gen Psychiatry 60:49–56

Hoffmann ETA (1967) The best tales of Hoffmann. Dover, New York

Hughlings JJ (1932) In: Taylor J (ed) Selected writings of John Hughlings Jackson. London, Hodden and Stoughton

Husserl E (1959) Erste Philosophie 1923/24. II. Theorie der Phaenomenologischen Reduktion. Martinus Nijhoff, The Haag

Husserl E (1960) Cartesian Meditations: An Introduction to Phenomenology. Cairns, D., trans. The Hague, Martinus Nijhoff. Originally published in 1931

Husserl E (1966) Analysen zur passiven Synthesis aus Forschungsmanuscripten 1918–1926. Martinus Nijhoff, The Hague

Husserl E (1973) Experience and Judgment, Investigations in a Geneology of Logic. Evanston, Ill:NorthwesternUniversity Press

Husserl E (1988) Cartesian meditations: an introduction to phenomenology. Martinus Nijhoff, The Hague, The Netherlands. Originally published in 1931

Husserl E (1997) Thing and space: lectures of 1907 (trans: Richard R). Kluwer, Dordrecht

Ichiro Yamaguchi (1982) *Passive Synthesis und Intersubjektivtät bei Edmund Husserl. The* Hague: Martinus Nijhoff

Ionasescu V (1960) Paroxysmal disorders of the body image in temporal lobe epilepsy. Acta Psychaitr Scand 35:171–181

Irwin HJ (1985) Flight of mind: a psychological study of the out-of-body experience. Scarecrow, Metuchen, NJ

James W (1961) The varieties of religious experience. 1902. Collier Macmillan, New York

Jantz H, Beringer K (1944) Das syndrom des schwebeerlebnisses unmittelbar nach kopfverletzungen. Nervenarzt 17:197–206

Jaspers K (1913) Über leibhaftige Bewusstheiten. Bewussheitstäuschungen., ein psychopathologisches Elementarsymptom. Z Psychopathol 2:150–161

Jaspers K (1946) Allgemeine psychopathologie. Allgemeine psychopathologie, 4th edition. Springer, Berlin. English edition: General psychopathology (1963) (trans: Hoenig J, Hamilton MW). Manchester University Press, Manchester, UK

Kafka F (1965) Diaries (1910–1913) (trans: Joseph K). Schocken Books, New York

Kafka F (1979) The basic Kafka. Pocket Books, New York

Kamiya S, Okamoto S (1982) Double consciousness in epileptics: a clinical picture and minor hemisphere specialization. In: Akimoto H, Kazamatsuri H, Seino M, Ward A (eds) Advances in epileptology, 13th epilepsy international symposium. Raven, New York, pp 397–401

Kurz G (1980) *Traum-Schrecken, Kafkas literarische Existenzanalyse*. Stuttgart: J.B. Metzlersche Verlag

Lacan J (1977) The mirror stage as formative of the function of the I as revealed in psychoanalytic experience. In: Sheridan A (trans) Écrits: a selection. W. W. Norton, New York

Legrand D, Newen A, Vogeley K (2003) How not to find the neural signature of self-consciousness. Conscious Cogn 12:544–548

Lhermitte J (1939) L'Image de Notre Corps. Nouvelle, Paris

Lhermitte J (1951) Visual hallucination of the self. Br Med J 1:431–434

Lhermitte F (1983) 'Utilization behaviour' and its relation to lesions of the frontal lobes *Brain* 106:237–255

Lishman JR, Lee DN (1973) The autonomy of visual kinaesthesis. Perception 2:287–294

Lopez C, Halje P, Blanke O (2008) Body ownership and embodiment: vestibular and multisensory mechanisms. Neurophysiologie Clinique/Clin Neurophysiol 38:149–161

Lukianowicz N (1958) Autoscopic phenomena. Arch Neurol Psychiat 80:199–220

Maravita A, Iriki A (2004) Tools for the body schema. Trends Cogn Sci 8:79–86

Menninger-Lerchenthal E (1935) Das Truggebilde der Eigenen Gestall. Karger, Berlin

Menninger-Lerchenthal E (1961) Heautoskopie. Wien Med Wochenschr 111:745–756

Merleau-Ponty M (1964) The child's relation with others. In: Edie JE (ed) The primacy of perception. Northwestern University Press, Evanston, IL, pp 96–155

Merleau-Ponty M (1968) The visible and the invisible (trans: Alphonso L). Northwestern University Press, Evanston

Merleau-Ponty M (1973) In: Claude L (ed) The prose of the world (trans: O'Neill J). Northwestern University Press, Evanston

Metzinger T (2003) Being no one. MIT Press, Cambridge

Metzinger T (2005) Out-of-body experiences as the origin of the concept of a "soul". Mind Matter 3:57–84

Metzinger T (2006) Précis: being no one. Psyche 12.4:1–7

Metzinger T (2006b) Reply to Gallagher: different conceptions of embodiment. Psyche 12:1–7

Miall RC (2003) Connecting mirror neurons and forward models. Neuroreport 14:2135–2137

Milner AD, Goodale MA (1995) The visual brain in action. Oxford University Press, Oxford

Mishara AL (1990) Husserl and Freud: time, memory and the unconscious. Husserl Stud 1990(7):29–58

Mishara AL (1994) A phenomenological critique of commonsensical assumptions of DSM-III-R: the avoidance of the patient's subjectivity. In: Sadler J, Schwartz M, Wiggins O (eds) Philosophical perspectives on psychiatric diagnostic classification. Johns Hopkins Series in Psychiatry and Neuroscience, Baltimore, pp 129–147

Mishara AL (1995) Narrative and psychotherapy: The phenomenology of healing. *American Journal of Psychotherapy* 49:180–195

Mishara AL (in press a) The 'Unconscious' in Paranoid Delusional Psychosis? Phenomenology, Neuroscience, Psychoanalysis. To appear in: Lohmar, D., Brudzinska, J. (eds.). *Founding Psychoanalysis*. New York: Springer

Mishara AL (in press b). Klaus Conrad (1905–1961): Why we need to 'understand' the patient's experience to study the neurobiology. *Schizophrenia Bulletin*

Mishara AL (in press c) Human ambivalence to body: Precondition for Social Cognition and its Disruption in Neuropsychiatric Disorders. *Philosophy, Psychiatry & Psychology*

Mishara AL (in press d) Kafka's doubles, paranoia and the brain: Hypnagogic vs. hyper-reflexive models of disruption of self in neuropsychiatric disorders and anomalous conscious states. *Philosophy, Ethics, and Humanities in Medicine (PEHM)*, PubMed Central Open Access Journal http://www.peh-med.com/

Mishara AL (1997) Binswanger and phenomenology. In: Embree L, Behnke EA, Carr D, Evans JC, Huertas-Jourda J, Kockelmans JJ, McKenna W, Mickunas A, Mohanty JN, Seebohm TM, Zaner RM (eds) Encyclopedia of phenomenology. Kluwer, The Hague, pp 62–66

Mishara AL (2001) On Wolfgang Blankenburg, common sense, and schizophrenia. Philos Psychiatry Psychol 8:317–322

Mishara AL (2004) The disconnection of external and internal in the conscious experience of schizophrenia: phenomenological, literary and neuroanatomical archaeologies of self. Philosophica 73:87–126

Mishara AL (2005) Body self and its narrative representation in schizophrenia: does the body schema concept help establish a core deficit? In: De Preester H, Knockaert V (eds) Body image and body schema. John Benjamins, Amsterdam, pp 127–152

Mishara AL (2007a) Is minimal self preserved in schizophrenia? A subcomponents view. Conscious Cogn 16:715–721

Mishara AL (2007b) Missing links in phenomenological clinical neuroscience? Why we are still not there yet. Curr Opin Psychiatry 20:559–569

Mishara AL, Schwartz M (1997) Psychopathology in the light of emergent trends in the philosophy of consciousness, neuropsychiatry and phenomenology. Curr Opin Psychiatry 10:383–389

Nadler JW, Angelaki DE, DeAngelis GC (2008) A neural representation of depth from motion parallax in macaque visual cortex. Nature 452:642–655

Navon D (2001) The puzzle of mirror reversal: A view from clockland Psycoloquy 12:1–27

Nowak LG, Bullier J (1997) The timing of information transfer in the visual system. Extrastriate Visual Cortex Primates 12:205–241

Paillard J (1991) Knowing where knowing how to get there. In: Paillard J (ed) Brain and Space. Oxford University Press, Oxford, pp 461–481

Paillard J (1991a) Knowing where knowing how to get there. In: Paillard J (ed) Brain and Space. Oxford University Press, Oxford, pp 461–481

Paillard J (1991b) Motor and representational framing of space. In: Paillard J (1991) Knowing where knowing how to get there. In: Paillard J (ed) Brain and Space. Oxford University Press, Oxford, pp 163–182

Paillard J (1999) Body schema and body image: d double dissociation in deafferented patients. In: Gantchev GN, Mori S, Massion J (eds) Motor control, today and tomorrow. Academic Publishing House, Sophia, pp 197–214

Paillard J (2005) Vectorial versus configural encoding of body space: a neural basis for a distinction between body schema and body image. In: De Preester H, Knockaert V (eds) Body image and body schema, interdisciplinary perspectives on the body. John Benjamins, Amsterdam, pp 89–109

Plessner H (1975) Die Stufen des Organischen und der Mensch. de Gruyter, Berlin/New York

Richter JP (1959) Siebenkas. Carl Hansen Verlag, Werke, Munchen (original, 1796)

Ruby P, Decety J (2001) Effect of subjective perspective taking during simulation of action: a PET investigation of agency. Nat Neurosci 4:546–550

Sartre JP (1966) In: Being and nothingness. A phenomenological essay on ontology (trans: Barnes HE). Simon & Schuster, New York (Original, 1943)

Sedman G (1966) A comparative study of pseudohallucinations, imagery and true hallucinations. Br J Psychiatry 112:9–17

Seebohm T (1962) Die Bedingung der Moeglichkeit der Transcendental Philosophie. Bouvier Verlag, Bonn

Seebohm T (1989) Transcendental phenomenology. In: Mohanty JN, McKenna W (eds) Husserl's phenomenology: a textbook. CARP and University Press of America, Washington DC, pp 345–385

Silberer H (1909) Bericht über eine Methode, gewisse symbolische Halluzionations-Erscheinungen hervorzurufen und zu beobachten. Jahrbuch für psychoanalytische und psychopathlogische Forschung 3:661–723

Sollier P (1903) Les Phénomènes d'Autoscopie. Félix Alcan, Paris

Spence SA, Brooks DJ, Hirsch SR, Liddle PF, Meehan J, Grasby PM (1997) A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control) Brain 120:1997–2011

Stawarska B (2004) The body, the mirror and the other in Merleau-Ponty and Sartre. In: Gallagher S, Watson S (eds) Ipseity and alterity: interdisciplinary approaches to intersubjectivity. Publications de l'Université de Rouen, Rouen, pp 175–186

Stephen JM, Aine CJ, Christner RF, Ranken D, Huang M, Best E (2002) Central versus peripheral visual field stimulation results in timing differences in dorsal stream sources as measured with MEG. Vision Research 42:3059–3074

Tadokoro Y, Oshima T, Kanemoto K (2006) Postictal autoscopy in a patient with partial epilepsy. Epilepsy Behav 9:535–540

Todd J, Dewhurst K (1955) The double: its psycho-pathology and psychophysiology. J Nerv Ment Dis 122:47–55

Twemlow SW, Gabbard GO, Jones FC (1982) The out-of-body experience: A phenomenological typology based on questionnaire responses. American Journal of Psychiatry 139: 450-455

Uhlhaas PJ, Mishara AL (2007) Perceptual anomalies in schizophrenia: integrating phenomenology and cognitive neuroscience. Schizophr Bull 33:142–156

Vogeley K, Bussfeld P, Newen A, Herrmann S, Happé F, Falkai P, Maier W, Zilles K (2001) Mind reading: neural mechanisms of theory of mind and self-perspective. NeuroImage 14:170–181

Vogeley K, May M, Ritzl A, Falkai P, Zilles K, Fink GR (2004) Neural correlates of first-person perspective as one constituent of human self-consciousness. J Cogn Neurosci 16:817–827

Weizsäcker V. von (1997) Der Gestaltkreis, dargestellt als physiologische Analyse des optischen Drehversuchs. Original 1933. In: Viktor von W (ed) Gesammelte Schriften, Bd. 4, (S. 23–61). Suhrkamp, Frankfurt a.M.

Weizsaecker V. von (1950) Der Gestaltkreis. Theorie der Einheit von Wahrnnehmung und Bewegung 4th Aufl. Georg Thieme Verlag, Stuttgart

Wigan AL (1844) A new view of insanity: the duality of mind. Longman, Brown, Green & Longmans, London

Wiggins OP (1994) Commentary on self-consciousness, mental agency, and the clinical psychopathology of thought insertion. Philos Psychiatr Psychol 1:11–12

Wiggins OP, Spitzer M (1997) Cognitive science. In: Embree L et al. (ed) Encyclopedia of phenomenology. Kluwer, The Hague, The Netherlands, pp 101–104

Wisdom JO (1953) The concept of "phantom-body". *Proceedings of the XIth International Congress of Philosophy: Vol VII: Philosophical psychology* (pp. 175–179). Amsterdam: North Holland Publishing Co

Wolpert DM, Doya K, Kawato M (2003) A unifying computational framework for motor control and social interaction. Philos Trans R Soc B: Biol Sci 358:593–602

Zutt J (1953) Aussersichsein und auf sich selbst zurückblicken als ausnahmezustand. Nervenarzt 24:24–30

Galati G, Committeri G, Sanes JN, Pizzamiglio L (2001) Spatial coding of visual and somatic sensory information in body-centred coordinates. Eur J Neurosci 14.4:737–746

Phenomenology as Description and as Explanation: The Case of Schizophrenia

Louis A. Sass

Introduction

The phenomenological approach in both philosophy and psychiatry has often been characterized as a descriptive rather than an explanatory enterprise. One can understand this statement in various ways. The general idea, however, is that the purpose of phenomenology is to describe and define the nature and varieties of human experience rather than to give an account of the causal mechanisms or efficacious processes that bring it about.

At an early phase of his work, Edmund Husserl (1859–1938) – phenomenology's founder – did indeed present phenomenology as a purely descriptive approach that excludes all concern with both genesis and causation (Bernet et al. 1993, p. 195). In the classic preface to his *Phenomenology of Perception*, Maurice Merleau-Ponty (1962, pp. vii–viii) characterizes phenomenology as "a matter of describing, not of explaining or analyzing … (as an attempt) to give a direct description of experience as it is without taking account of its psychological origin and the causal explanations which the scientist, the historian, or the sociologist may be able to provide." Similar views are common in phenomenological psychiatry and psychopathology. The phenomenological psychiatrist, Wolfgang Blankenburg (1971/1991, p. 4, 27), e.g., explicitly denies that his account of the "basic disorder" (*Grundstörung*) in schizophrenia is intended to have any etiologic significance; he aims, he says, only to capture the "essence" of typically schizophrenic abnormalities (see also Buytendijk 1987, p. 130).

The close connection between phenomenology and description can hardly be disputed. Indeed, as Heidegger (1962) points out in *Being and Time*, the phrase "'descriptive phenomenology' … is at bottom tautological" (p. 59). The identification of phenomenology with description *alone* – not to mention the very distinction between description and explanation itself[1] – is not, however, nearly as straightforward or as

L.A. Sass
Department of Clinical Psychology, Rutgers University

---

[1] See, e.g., Simon (2000, p. 25): "The line between descriptive and explanatory laws is not a sharp one, for we may find all kinds of intermediate cases – especially for qualitative explanations." As Simon points out, both "causal" and "explanation" are terms "gravid with implications" and highly "problematic" (p. 22). See Michotte (1963) for an experimental demonstration of the difficulty of separating perceptual observation from causal attribution.

universally accepted as it may first appear. Careful review of the phenomenological tradition, together with reflection on the meaning of the concept "explanation", suggests, in fact, that there are a number of crucial ways in which phenomenology can indeed play an explanatory role. The purpose of the present chapter is to offer a clear and reasonably succinct, contemporary overview of these complex issues for use by psychiatrists, clinical psychologists, and other students of psychopathology.[2]

Incidentally, the phenomenologists who eschew explanatory ambitions should not necessarily be understood as arguing for the causal irrelevance or causal independence of conscious experience: typically, they mean to imply a bracketing or setting-aside of all such questions in order to facilitate a purified description of subjective life. Also, in this paper, the term "phenomenology" is used in the standard philosophical and continental sense: that is, to refer to the study of lived experience and of how things manifest themselves to us within and through such experience (Moran 2000; Sokolowski 2000, p. 2). By contrast, in mainstream Anglophone psychiatry, the term "phenomenology" typically refers simply to signs and symptoms that can readily be observed.

Description and Explanation, Motivation and Causation

In the above-quoted preface to the *Phenomenology of Perception*, Merleau-Ponty (1962) goes on to point out that, over time, Husserl came to adopt a broader and more ambitious view of phenomenology, advocating the need to supplement "static" or "descriptive phenomenology" with phenomenology of a "genetic" or "constructive" type. Indeed, Husserl (1999) himself came to speak of "explanatory" phenomenology – a "phenomenology of regulated genesis" (p. 318). In his late work, Husserl (1989) also spoke of "motivational" relationships or even a "motivational causality" whose study clearly fell within the province of phenomenology (p. 227; 1999, p. 320). He described motivation as providing the "fundamental lawfulness of spiritual life" (1989, pp. 231, 241f).[3]

Husserl carefully distinguished motivational causality from causality as understood in a narrower sense – that is, from the "natural causality" or "real causality" of physical nature (which, presumably, involves the efficient form of causality). "Motivation," for Husserl, concerns the attitude and orientation of the subject. It is unlike blind causality, for it involves the subject's viewpoint *on* or interpretation *of* the world. It is unlike processes of reasoning, for it has a more spontaneous, immediate,

---

[2] The present chapter overlaps considerably with another, rather longer paper: Sass and Parnas (2007).

[3] It is not easy to provide a succinct *précis* of Husserl's complex views on genetic phenomenology and motivational relationships. On these difficulties, see Bernet et al. (1993, p. 196). For attempts to clarify these issues, see Steinbock (1995), Depraz (2001).

and implicit quality. Motive, according to Merleau-Ponty (1962, pp. 49f) as well as Husserl, must be distinguished from both reason and cause: it is "one of those 'fluid' concepts that have to be formed if we want to get back to the phenomena. One phenomenon releases another, not by some objective efficacy, like that which links events in nature, but by the meaning it offers – there is a *raison d'etre* that orients the flux of phenomena without being explicitly posited in any one of them, a sort of operant reason" (Merleau-Ponty quoted in Carman 2005, p. 84). Edith Stein, who studied under Husserl, defined motivation as involving a relation or connection between experiences and their antecedents in which one finds "an *arising* of the one *from* the other, an effecting or being effected of one *on the basis* of the other, *for the sake of* the other" (Stein, quoted in Wrathall 2005, p. 116).

Husserl's concept of motivation is broader than the ordinary concept of motive, disposition, or ground for action. It covers many forms of implicative interdependence between mental acts and experiences that contribute to the coherence and unity of consciousness, both in its synchronic and diachronic aspects (1989, pp. 223–293).[4]

It is clear, in any case, that Husserl (1989, p. 402) gradually moved away from Wilhelm Dilthey's sharp opposition between description (as the goal of the human sciences) and explanation (as the goal of the natural sciences). In his lectures on phenomenological psychology of 1925, Husserl (1977) spoke of "ultimate unclarities concerning the mutual relation of nature and mind and of all the sciences which belong to these two titles … what seems at first obviously separated, upon closer inspection turns out to be obscurely intertwined, permeating each other in a manner very difficult to understand" (p. 39).

The concepts of both "explanation" and "causation" are problematic, heterogeneous, and "obscurely intertwined." Both concepts have been disputed since ancient times and continue to be highly contested in contemporary philosophy (Audi 1999, p. 127; Crane 1995). The *Cambridge Dictionary of Philosophy* defines "explanation" in simple and general terms: as "an act of making something intelligible or understandable, as when we explain an event by showing how or why it occurred" (Audi 1999, p. 298). The paradigm cases of explanation typically refer to efficient causal mechanisms or processes in the physical world. There are, however, also concepts of both "motivational" and "mental" causation (Husserl 1989; Heil and Mele 1993). Indeed, "causal relevance" can be defined quite broadly – as requiring

---

[4] Motivation can operate through associative and other links among the *contents* of awareness. These would have some analogies with the kinds of links emphasized by analytic philosophers who speak of "mental causation" or the "practical syllogism" – see Section on Explanatory Relevance of the Mental or Subjective Domain: Preliminary Considerations below. Motivation (in Husserl's sense) can, however, also operate through formal or structural aspects of the (noetic) *act* of consciousness itself – as, e.g., when inner time consciousness serves as a necessary condition for the unity of the flux of experiences, or when a distorted mode of self-experience is expressed in specific kinds of delusional beliefs (see Section on Expressive Relationships below) (Husserl 1989, p. 238).

only that a given attribute or factor (the cause) "makes a difference" to the probability of the occurrence of a given property (the effect) (Grunbaum 1993, p. 163). Further, the possible forms of explanation need not be restricted to causation alone: they can also involve other forms of relationship that reveal the underlying unity or interdependence of a group of phenomena. Indeed, explanatory factors can be said to cover "*all* those things to which any event or process can be ascribed, *anything* in the light of which it can be said to make sense" (Lawson-Tancred 1995, pp. 418–419, emphasis added). Explanation can be defined as "an apparently successful attempt to increase the understanding of (a given) phenomenon" (Wilson and Keil 2000, p. 89).

Given the problematic nature of these issues, it is hardly surprising there is no consensus about or clear statement of these issues in psychopathology or the more specific domain of phenomenological psychopathology. In this chapter, I do not claim to offer anything approaching an all-inclusive or fully rigorous classification of the forms of explanation – which have been disputed at least since Aristotle. I only wish to indicate certain important forms of explanation relevant to psychopathology that are neglected in psychiatry, forms that refute the widespread assumption that phenomenological accounts are unimportant because they are "*merely*" descriptive in nature.

I will provide a preliminary taxonomy of six forms of phenomenological explanation, each of which fits into one of two general explanatory perspectives. These general perspectives are distinguished according to whether the relationships at issue apply to phenomena that occur simultaneously or in succession. Whereas the first explanatory perspective involves what might be called phenomenological *implication*, the second has a causal or at least quasi-causal significance.[5]

In this article, I will focus on schizophrenia, and will use my own, phenomenological account of the disorder (developed together with Josef Parnas; see Sass and Parnas 2003) to illustrate the above-mentioned issues. I believe, however, that most of my formulations in this paper have general application to other forms of psychopathology. The main purpose of this article is to *use* this contemporary account of schizophrenia as a way of illustrating the explanatory relevance a phenomenological approach can have. The account, which has strong affinities with the work of several other phenomenological psychopathologists (especially Minkowski and Blankenburg; Sass 2001a), has been developed and defended in detail elsewhere (Parnas 2000, 2003; Sass 1992a, 1994, 1998a, 2003a, 2003b; Sass and Parnas 2003). Here I sketch it as briefly as possible, hoping only to indicate its relevance to the broad range of schizophrenic symptoms before moving on to consider a variety of distinct ways in which it may have more than merely descriptive significance. My concern here is not to prove the *correctness* of this particular interpretation of schizophrenia, but only to lay out a set of explanatory *possibilities*.

---

[5] I am aware that, according to many philosophers, a cause can be simultaneous with its effect (Mackie 1974, p. 161). A future, more fully adequate taxonomy will doubtless need to take this into account.

Disturbed Ipseity: A Phenomenolgical Account of Schizophrenia

The purpose of a phenomenological investigation is to give an accurate account of the form and structure of subjective life. The term "account" is appropriately vague in this context, since it can refer to both description and explanation.

Phenomenology is concerned both with the ways objects of awareness *are given* in experience (Husserl called these *noematic* aspects, which include the experiential content and mode) and also with the nature of the acts of awareness by which these objects and modes *are formed* (constituted) in the intentional stream of awareness (the so-called *noetic* aspects). (In this chapter, the terms "intentional" and 'intentionality' refer, not to the issue of volition, but to the essential "aboutness" of conscious awareness or mental life – to the fact that mental acts always have or are directed toward some object. For introductions to phenomenology, see Moran (2000), Sokolowski (2000); re phenomenological psychopathology, see Parnas and Zahavi (2002), Sass (1992a, b)).

According to the view to be presented here, the core abnormality in schizophrenia is a particular kind of disturbance of consciousness and, especially, of the sense of self or ipseity that is normally implicit in each act of awareness. (*Ipseity* derives from *ipse*, Latin for "self" or "itself." Ipse-identity or ipseity refers to a crucial sense of self-sameness, of existing as a subject of experience that is at one with itself at any given moment (Ricoeur 1992; Henry 1973; Zahavi 1999)). This self or *ipseity disturbance* has two main aspects or features that may at first sound mutually contradictory, but are in fact complementary. The first is hyperreflexivity – which refers to a kind of exaggerated self-consciousness, that is, a tendency for focal, objectifying attention to be directed toward processes and phenomena that would normally be "inhabited" or experienced as part of oneself. The second is diminished self-affection – which refers to a decline in the (passively or automatically) experienced sense of existing as a living and unified subject of awareness. (Please note: the term "affection" refers to a process of being affected by something; it has nothing to do with the notion of fondness, or liking of oneself.)[6] This two-faced disturbance of ipseity disrupts the normal pre-reflective sense of existing as a self-presence that is the "I-center" or "central point of psychic life" – what, in Husserlian phenomenology, could be called the "source-point of the rays of attention," "center of reception," or "pole of the affections" (Bernet et al. 1993, pp. 209ff).

These mutations of the act of awareness are typically, perhaps necessarily, accompanied by alteration in the objects or field of awareness – namely, by disruption of the focus or salience with which objects and meanings emerge from a background context; I refer to the latter alteration as disturbed perceptual or conceptual "grip" or "hold" on the world (Merleau-Ponty 1962, p. 240; Dreyfus 2002).

---

[6] The term "affection" in the phrase "self-affection" is intended to suggest a quality of passivity and also affinity with the affects; this contrasts with mental processes or events that have a more active and purely cognitive nature. To be affected by something means to be touched, moved, or motivated by it – a process that is primordially linked to emotionality (Henry 1973; Parnas 2003).

Table 1 A phenomenological account of schizophrenia

The central feature, Ipseity disturbance (a disruption of consciousness and self-experience), has
   three facets:
Noetic abnormalities (abnormalities in the constituting *act* of awareness)
   1. Hyperreflexivity
   2. Diminished self-affection
Noematic abnormality (abnormalities in the constituted *objects* or *field* of awareness –
   perceptual or cognitive)
   3. Disturbed grip or hold

Merleau-Ponty (1962) describes this grip or hold as involving "a certain culmination
and optimum balance in the perceptual process," in a perceptual "field in which
richness and clarity are in inverse proportion to each other" (p. 318), thus providing
"a spectacle as varied and as clearly articulated as possible" (p. 250).[7] He states that
this "maximum sharpness of perception and action points clearly to a perceptual
ground, a basis of my life, a general setting in which my body can co-exist with the
world" (p. 250); this, in turn, requires normal self-affection and an appropriate bal-
ance between tacit and explicit modes of awareness.

   These descriptions of hyperreflexivity and diminished self-affection, on the one
hand, and of loss of perceptual/conceptual "hold," on the other, are attempts to
characterize, respectively, the noetic and the noematic infrastructures of the schizo-
phrenia patient's characteristic mode of experience and lifeworld (see Table 1).

   It should be noted that the hyperreflexivity in question is not, at its core, an intel-
lectual, volitional, or "reflective" kind of self-consciousness. Most basic to schizo-
phrenia is a kind of "operative" hyperreflexivity that occurs in an automatic fashion.
This has the effect of disrupting awareness and action by means of an automatic
popping-up or popping-out of phenomena and processes that would normally
remain in the tacit background of awareness (where they serve as a medium of
implicit self-affection), but that now come to be experienced in an objectified and
alienated manner (see Merleau-Ponty 1962, p. xviii, citing Husserl re. "operative
intentionality" – *fungierende Intentionalität*).[8]

   The self-disturbance being postulated is not fundamentally a disturbance of self-
image or social identity; nor does it primarily involve the continuity of identity over
time (which is not to say that these aspects of selfhood will not be affected in any
way). It pertains, rather, to a more fundamental sense of existing as an experiencing
entity of some kind, as a kind of implicit subject-pole that would normally serve as
the vital center-point of subjective life. This fundamental feature of normal

---

[7] Maximal "grip" or "hold" on the world requires a "certain balance between the inner and outer
horizon." If seen from too close, a living body, now divorced from its background, can seem an
outlandish "mass of matter"; if seen from too far away, it may lose its "living value" and appear
as a puppet or automaton (Merleau-Ponty 1962, p. 302).

[8] Contrast this with the kind of awareness we normally have of our bodies: "not … knowledge in
thematized form. [Rather] an inarticulate and indistinct familiarity completely devoid of positional
and disclosing consciousness" (Gurwitsch 1964, p. 302, describing Merleau-Ponty's views).

awareness – known as self-affection – is especially difficult to articulate in rich descriptive detail precisely because it is such a pervasive, fundamental, and obvious aspect of consciousness. The writer Antonin Artaud (who suffered from schizophrenia) was referring to an aspect of this when he spoke of what he called "the essential illumination" and this "phosphorescent point," equating this illuminating center-point with the "very substance of what is called the soul," and describing it as a prerequisite for avoiding "constant leakage of the normal level of reality" (Artaud 1976, pp. 169, 82; 1965, p. 20). Another patient with schizophrenia described the condition of lacking this crucial if ineffable self-affection that is essential to normal ipseity: "I was simply there, only in that place, but without being present" (Blankenburg 1971/1991, pp. 42, 77).

Finally, the mutations of the perceptual or cognitive *field* of awareness (disturbed grip or hold) involve not just any kind of obscurity or disorganization. As we shall see, there are forms of confusion or "perplexity" (*Ratlosigkeit*; Störring 1987), highly characteristic of schizophrenia, that derive from an absence of vital, motivating concerns (which is a concomitant of normal self-affection), and from an emergence into awareness of what would normally have been too self-evident to be noticed.

The two-faceted disturbance of the act of awareness (hyperreflexivity and diminished self-affection) can be shown to be implicated in each of the three major syndromes of schizophrenia recognized in contemporary research: the "positive," "disorganization," and "negative" syndromes (see Sass 2003; Sass and Parnas 2003).

Explanatory Relevance of the Mental or Subjective Domain: Preliminary Considerations

Before I attempt to lay out the various types of phenomenological explanation, it will help to situate the phenomenological approach in relation to certain traditional as well as contemporary Anglo-American conceptions of the nature of explanation and understanding in psychology and psychopathology.

Traditionally, a distinction has often been made between "explanation" and "understanding," with explanation being said to pertain to (causally determined) physical processes, and understanding to be appropriate for the comprehension of human experience, action, and expression (von Wright 1971). Human actions and experiences have been assumed to be recalcitrant to causal explanation for at least two reasons: first, because they have a particularistic, context-embedded quality that defies the possibility of theoretical generalization; and second, because mechanistic or deterministic causal models are presumed to be inapplicable to the realm of goal-directed activity, which is dominated by motive or reason. In recent years, however, aspects of this traditional dichotomy have been questioned by analytic philosophers who argue (against certain followers of Wittgenstein or of hermeneutics, such as Paul Ricoeur) that reasons *are* in fact a species of cause. The influential work of Donald Davidson (1980) calls attention to a perhaps intuitively obvious

fact: namely that a reason *does* play a role in a causal account of a given action when it is claimed to be *the* reason why an agent actually did act as he did. In such a case, however, the reason (according to Davidson) must fulfill certain logical requirements such that it can be schematized as part of what is called a "practical syllogism."[9]

On the account offered by Davidson and other analytic philosophers (the belief-desire-intention paradigm; see Cummins 2000, p. 127), mental causality involves a triangulation of desire, belief, and a dispositional belief system. If I desire to buy an umbrella and if I also believe there is a store across the street that carries umbrellas, and if, in addition, my desire or belief is not *in*congruent with other beliefs (e.g., I believe I can cross the street, etc), then this triangular interaction has a possible causal role in explaining my finally crossing the street to buy the umbrella. On this sort of account, the ensuing intentional action is like the conclusion of a syllogism (the "practical syllogism"): it follows from it logically. "For a desire and a belief to explain an action in the right way," Davidson writes, "they must cause it in the right way, perhaps through a chain or process of reasoning that meets standards of rationality" (1980, p. 232f). Here a desire, motive, or reason (I want to buy an umbrella) is re-described as something that has real causal efficacy.[10] Many contemporary Anglo-American philosophers and cognitive scientists assume that this is the *only* way in which mental events could have a causal role; they present it as the *only* alternative to an account in terms of physical causation (but see Griffiths 1997, p. 244).[11] It is reasonable to ask, therefore, about the relationship between this sort of account and that which phenomenology has to offer.

It should be evident from the above that discussion of mental causation in recent analytic philosophy has focused largely on the question of the rational coherence and potential explanatory significance of individual mental *contents* – e.g., the belief *that* there is a store across the street; the desire *that* one buy an umbrella. Only phenomena that can be said to contain (or to be describable in terms of) this sort of "propositional content" are capable of serving the kind of rationalizing or justificatory function that is required by the practical syllogism (Evnine 1991, p. 11). It is here that the distinctness of the phenomenological perspective becomes important.

---

[9] For critiques of Davidson, see (Evnine 1991; Sass 2001b, pp. 264–274).

[10] Actually, Davidson's position on the causal efficacy of mental contents is very difficult to pin down, as various commentators have remarked (see Sass 2001b, pp. 287–290, for discussion and various references, including Kim 1985). Wakefield and Eagle (1997) offer a clear example of a reading that interprets Davidson as ascribing real causal efficacy to "mental representations" existing "in the head" (p. 323). Sass (2001b) criticizes the coherence of Davidson's position and questions its actual relevance for psychological explanation.

[11] Griffiths (1997, p. 244) argues that "many problems in the philosophy of mind have been occasioned by the loss of … flexibility in our thought about mental contents" that was occasioned by adoption of the philosophical "propositional attitude" theory. As Griffiths points out, the latter approach (exemplified by Donald Davidson) derives from Aristotle's formalized model of action explanation via the "practical syllogism."

Phenomenology does not ignore the content of experience; nor does it deny that some aspects of this may be analyzable in terms of sentence-like propositional attitudes. But the emphasis of a phenomenological account, the focus of its description effort, is directed elsewhere – toward formal or structural features that involve more pervasive aspects or *infrastructures* of human experience (e.g., modes of temporal or spatial experience, general qualities of the object world, forms of self-experience).[12] These latter have more in common with the phenomena of mood or cognitive style than they do with particular beliefs, perceptual contents, or wishes whose significance could be captured as a sentence or a logical proposition.[13]

Consider, e.g., the concepts of diminished self-affection, hyperreflexivity, and loss of cognitive-perceptual "hold": these are not reasons nor are they causes, at least of a physicalistic kind. In what sense, then, can these three concepts be said to have any *explanatory* significance or to contribute to a genetic or *causal* explanation of schizophrenia? Pursuing these questions will lead in two directions: first toward the question of the relationship *between* the three just-mentioned aspects; and second, toward an examination of the roles these aspects may play in determining both short- and long-term developmental transformations of schizophrenia.

In the following discussion of the explanatory relevance of a phenomenological account, it will be useful to distinguish two general, explanatory perspectives according to whether the relationships to be described are primarily synchronic or diachronic – that is, whether they apply to phenomena viewed as occurring simultaneously or in succession.[14] First I consider the domain of the synchronic, where I distinguish three kinds of relationship: *equiprimordial*, *constitutive*, and *expressive*. Although these do not involve either causation or genesis over time, they do involve forms of what might be called "phenomenological *implication*" – and thus they perform an explanatory rather than merely descriptive function. Later I turn to the domain of the diachronic. Here I will distinguish *primary*, *consequential*, and *compensatory* processes. All three help to account for the genesis of schizophrenic phenomena, and are potentially relevant to what might broadly be defined as *causal* accounts of the development of schizophrenic symptoms (see Table 2).[15]

---

[12] Jaspers (1963, p. 59): "… from the phenomenological point of view, it is only the form that interests us."

[13] See Sass (1998b) for an introduction to hermeneutic phenomenology, where the emphasis is on background or "horizonal" aspects of existence.

[14] This distinction between synchronic and diachronic dimensions is a simplifying abstraction, not meant to be taken too literally. All conscious processes are, in some sense, intrinsically temporal in nature.

There is at least a rough correspondence between my synchronic-diachronic distinction and Husserl's distinction between static and genetic phenomenology (see Husserl 1999, pp. 144, 319).

Philosophers have debated the question of the relationship between causation and temporal sequence, with some pointing out that a cause can sometimes be simultaneous with its effect. In this paper, however, I focus on possible causal *sequences*.

[15] Obviously, I am not using "causal" here in the narrow sense of mechanical efficient causality.

Table 2 Forms of phenomenological explanation

Synchronic relationships: phenomenological implication
    -Equiprimordial
    -Constitutive
    -Expressive relationships
Diachronic dimension: phenomenological causality
    -Primary
    -Consequential
    -Compensatory processes

Synchronic Relationships

Equiprimordial Relationships[16]

As I have said, *hyperreflexivity* and *diminished self-affection* involve fundamental distortions of the act of awareness: altered interplay of tacit and explicit components, and a concomitant loss of a grounding sense of existing as a subject of action and awareness. But how, one may ask, is one to understand the relationship *between* hyperreflexivity and diminished self-affection? One possibility is to view them as intimately intertwined yet distinct processes that can interact with or even give rise to each other. This process could work in either direction.

Careful phenomenological investigation suggests, however, that hyperreflexivity and diminished self-affection may, in many cases, *not* best be conceived as outcomes or indices of distinct processes but, rather, as aspects of a single whole that we simply happen to be describing from two different angles of vision. Indeed, it might be argued that these two disturbances are really one and the same phenomenon, the very *same* distortion of consciousness or subjectivity that we are merely describing in different words. Whereas the notion of hyperreflexivity emphasizes the way in which something normally tacit becomes focal and explicit, the notion of diminished self-affection emphasizes a complementary aspect of this process, the fact that what once was tacit is no longer being inhabited as a medium of taken-forgranted selfhood. Thus neither is more basic than the other; they are *equiprimordial* aspects of a fundamental (noetic) disturbance of the act of awareness. A clear theoretical grounding for this view is provided by the philosopher Michael Polanyi's (1964, 1967) account of the vector of conscious awareness (what Merleau-Ponty 1962, pp. 136, 157, called the "intentional arc") as a continuum stretching between the object of awareness (what he calls the "distal" pole), which is known in a focal or explicit way, and that which exists in the "tacit dimension," i.e., which is experienced in what Polanyi terms a more subsidiary, implicit, or tacit manner. A tacit or subsidiary awareness of kinesthetic and proprioceptive sensations serves

---

[16] The term "equiprimordial" is taken from Heidegger (1962).

as the very medium of pre-reflective selfhood, ipseity, or self-awareness (the "proximal" pole of the vector of awareness), which, in turn, is the medium through which all intentional activity is realized.

Constitutive Relationships

The relationship between this two-faceted noetic transformation of the *act* of awareness, on one hand, and the loss of perceptual/conceptual "hold" (a transformation of the noematic object or field), on the other, would also be misunderstood or oversimplified if it were conceived on the model of contingent causation between independent processes. I do conceive of the noetic aspects of the dissolution of intentionality (hyperreflexivity and diminished self-affection) as the more fundamental or *constitutive* disturbance; they are aspects of the *act* of consciousness whereby experience is constituted. In this view, notions like "dissolution of natural experience" (Binswanger), "loss of natural self-evidence" (Blankenburg 1971/1991), loss of "perspectival abridgment" (Sass, 1992a, chap. 4), and "loss of hold or grip" are alternative ways of describing the noematic or constitu*ted* aspects of this dissolution – which contribute to the peculiar "perplexity" (Störring 1987) that is so characteristic of schizophrenia.[17]

It is important to remember that normal ipseity, with its usual self-affection and balance between the tacit and the focal, is not only a condition for the experience of appetite and vital energy. It also provides a point of orientation: it is what grounds human motivation and organizes our experiential world in accordance with needs and wishes, thereby giving objects their "affordances," their significance for us as obstacles, tools, objects of desire, and the like. In the absence of this vital yet implicit self-affection, and the lines of orientation it establishes, the structured nature of the worlds of both thought and perception will be altered or even dissolved. For then there can no longer be any clear differentiation of means from goal; any reason for certain objects to show up in the focus of awareness while others recede; or any reason for attention to be directed outward toward the world rather than inward toward one's own body or processes of thinking. Without normal

---

[17] In this respect, I follow Husserl rather than Heidegger. Heidegger conceived of human existence as a condition of *being there* (*Dasein*) and questioned what he saw as his mentor, Husserl's, overly subjectivist and Cartesian conception of mind as *constituting* the experiential world. Although Husserl fully recognized there is no *noesis* (act of consciousness) without a correlative *noema* (object of consciousness), he nevertheless gives a special status to the noetic acts, which he describes as "animating construals" or "apprehensions" that are responsible for the transcendental constituting of the objects and field of our awareness (Husserl 1983, pp. 226, 238, 277). One may certainly debate the merits of a Heideggerian versus a Husserlian approach (Tatossian 1997, p.12). It is worth noting, however, that the Husserlian interest in constituting mental processes and the genesis of experiential worlds is more obviously congruent with the aspirations of contemporary psychology and cognitive science, which seek to identify mental processes that underlie and in this sense account for the experiential abnormalities.

self-affection, the world will be stripped of all the affordances and vectors of concern by which the fabric of normal, common-sense reality is knitted together into an organized and meaningful whole.[18]

In emphasizing the foundational role of hyperreflexivity and diminished self-affection, I am not suggesting that they exist independently of or prior to the noematic disturbance: they are not the cause but the *condition of possibility* for the disturbance of cognitive-perceptual hold.

This *constitutive* type of relationship does not, incidentally, conform to either of the two types of explanation that are countenanced by many recent Anglo-American philosophers: it is neither "a psychophysical link holding between states of affairs or events" nor "a relationship of making intelligible holding between sentences" (Taylor 1993, p. 326) – the latter being the only form of *mental* causation accepted by many analytic philosophers (Heil and Mele 1993; Sass 2001b). But, as the philosopher Charles Taylor (1993, p. 326) rightly notes, these two alternatives do not, in fact, exhaust the space of possibilities. Another possibility Taylor mentions is the "world-shaping relation" between the lived-body or corporeal subject and the world of experience – also an instance of a *constitutive* relationship.[19]

Expressive Relationships

A third type of relationship of mutual phenomenological implication – the *expressive* type – involves situations in which the (noematic) *content* of mental life seems to represent or express, in a more specific way, what appear to be more *general* formal or structural characteristics of mental life. (This distinction corresponds, in Heidegger's system, to the difference between "ontic" facts and "ontological" dimensions of existence; see Sass 1992b.) Take, for example, a delusion about dissolving, being controlled by an influencing machine, or being constantly recorded by videocameras. This sort of delusion may be understandable, not because it plays a role in a logical syllogism, but because it actually *expresses* or *emblematizes*, in relatively concrete form, more general or formal features of the prevailing state or mode of consciousness – in this case, the general state of ipseity-disturbance (see Merleau-Ponty's notion of an "emblem of being": Merleau-Ponty 1968, p. 270; Dreyfus and Wakefield 1988, p. 280).

Apparent logical contradictions in the content of a person's thoughts may become understandable in this way. Consider, for instance, the famous influencing-machine delusion of the patient Natalija (Tausk 1933) – a delusion that implies that Natalija experiences herself as, at the same time, godlike (at the center of the world, with all other entities existing only for her), but also a mere passive entity within

---

[18] See Dworkin et al (1998, pp. 390, 412) re role of the individual's "concerns" in determining the emotional meaning and general significance of events.

[19] Other types also fall outside this dualism of explanatory types; see below.

the world (a machine manipulated by others). The fact that the self can be experienced *both* as a passive mechanism *and* as a kind of solipsistic deity – sometimes at the same moment – can be understood if one recognizes that both these forms of self-experience are implicit in a hyperreflexive focus on the functioning of one's own mind and its role in the constitution of the experiential world (see Sass 1998a, b; Bovet and Parnas 1993).

Here phenomenology clearly plays more than a merely descriptive role. To articulate such expressive relationships provides an integrating vision, an understanding not of patterns of causal interaction but "of style, of logical implication, of meaning and value" (Geertz 1973, p. 145); and this serves an explanatory function (see also Minkowski, 1997, re. relationships of signification and expression).

Conclusion: Phenomenological Implication

On the present view, then, the three facets (hyperreflexivity, diminished self-affection, loss of hold), and also the form and emblematic content of experiential life, are linked together in relationships of necessary implication rather than contingent correlation or causal interaction. The implications in question are not, however, *logical* (like the practical syllogism) but, rather, *phenomenological* in nature, with the individual factors being understood as mutually implicative *aspects* or *expressions* of mental activity as a whole (Marbach 1993, p. 35). This is what Husserl was pointing to when he described "conscious life" as "contain(ing) an intentional inter-twining, motivation, mutual implication by meaning … which in its form and principle has no analogue at all in the physical" (Husserl 1977, p. 26). Similarly, Merleau-Ponty (1962) spoke of "internal links" between aspects of experience that "display one typical structure … standing in a relationship to each other of recipro-cal expression" (p. 157).[20] Phenomenological investigation is, in this way, less a matter of discovering interacting processes or of analyzing logical syllogisms than it is of *un-folding* the different facets of conscious life or activity in order to provide a richer grasp of its lived texture and internal structure.

The Diachronic Dimension

I turn now to the diachronic dimension, to questions concerning schizophrenia's devel-opment over time and the relative causal primacy of various kinds of processes. As we shall see, neither hyperreflexivity nor diminished self-affection is a singular or fully

---

[20] Merleau-Ponty (1962) uses the concept of "reciprocal expression" in a broad way – to refer to "internal links" between whole modes of experience: "Thus sexuality is not an autonomous cycle. It has internal links with the whole of active and cognitive being, these three sectors of behavior [sexuality, action, cognition] display one typical structure, and stand in a relationship to each other of reciprocal expression" (p. 157).

homogenous phenomenon; each can play what might be termed a *primary*, *consequential*, or *compensatory* role in the generation of experiential abnormalities. Although for brevity's sake I will focus on hyperreflexivity in the pages below, most of my points could also be developed with regard to the aspect of diminished self-affection.

Primary Hyperreflexivity

Both Husserl and Merleau-Ponty have described the multi-layered nature of human intentionality. The most fundamental level is an intentionality that can be described as "operative" or bodily in nature; this occurs on a *prereflective* level and is the medium in which habits and dispositions become sedimented (Merleau-Ponty 1962, p. xviii). Operative intentionality happens, phenomenologically speaking, in a passive, non-willed, or automatic manner.

What has just been said applies to *normal* human intentionality. Forms of *hyper-reflexive* intentionality can, however, also be distinguished according to whether they have an "operative" as opposed to a more "reflective" character. Here I use the label "*operative* hyperreflexivity" to denote a process afflicting the more fundamental levels of intentionality – a process in which the normally transparent field of experience becomes increasingly disrupted by unusual sensations, feelings, or thoughts that would normally remain in the background of awareness but that now pop into awareness and come to acquire object-like quality (a kind of spatialization of experience). Patient reports suggest that this is first experienced as a largely passive process, more like an affliction, typically involving cenesthesias, a loss of the automaticity of movement, and certain cognitive and perceptual disturbances – phenomena that, in the "basic-symptom" research, are designated with the apt term "basal irritation" (Klosterkötter et al. 1997). At the proximal pole of the vector of awareness, the same phenomenon manifests itself as a fundamentally diminished or altered sense of self-presence and presence to the world.

In its most primary form, then, this "irritation" may well occur in a largely passive manner, and therefore represents an "operative" rather than "reflective" kind of hyperreflexivity. This irritation may, in fact, be a rather direct consequence of a neurally based cognitive dysfunction. (The popping-up of normally tacit sensations could, e.g., result from disturbances of the hippocampus-based comparator system or from some other disturbance of "cognitive coordination" (Gray et al. 1991; Hemsley 2005; Phillips and Silverstein 2003) – to mention but two of the most plausible of current neurocognitive models.)

Consequential Hyperreflexivity

Primary "irritation" and ipseity disturbance do, however, attract further attention, thereby eliciting processes of scrutiny and self-exacerbating alienation ("*conse-quential* hyperreflexivity"). Although these may have a somewhat more active or

quasi-volitional quality (which is not to say, however, that they are *fully* conscious or volitional), they occur as byproducts of a more primary disturbance. A patient may, e.g., find himself paying more attention to odd kinesthetic sensations, or may find himself scrutinizing odd visual appearances in a way that only increases their oddness.

Thus a more primary ipseity-disturbance seems to allow, perhaps to inspire, a more reflective turning-inward and self-alienation of a mind that comes to take itself as its own object. And, in turn, this inwardness and self-reflection seems to contribute to an undermining (via externalization) of the normally tacit sources of self-affection.

Compensatory Hyperreflexivity

The primary disturbances of ipseity do not merely elicit fairly automatic consequences; they also inspire defensive or *compensatory* forms of hyperreflexivity that have a more goal-directed quality (without necessarily being conscious or fully volitional). Patients may attempt, for example, to reassert control and reestablish a sense of self by means of an introspective scrutinizing. Or they may engage in pseudo-obsessive intellectual ruminations in attempting to make up for a more primary sense of unnaturalness and unfamiliarity of the world and other people.

Diminished self-affection can also develop in a compensatory fashion. We know that, in Dissociative Identity Disorder and Post-Traumatic Stress Disorder, patients undergo a loss of the sense of their own reality or existence as experiential subjects that is, at least in part, defensively motivated. Similar developments can occur in schizophrenia-spectrum patients – who may have good reason to seek the escape inherent in self-obliteration.

Defensive or compensatory processes often have counterproductive effects, however. The more active or reflective forms of hyperreflexivity may, e.g., serve as the source of *further* alienations or diminishments of ipseity and perceptual meaning. This can be very disturbing to the patient. "My downfall was insight," explained one young man with schizophrenia, "too much insight can be very dangerous, because you can tear your mind apart." "Well look at the word 'analysis'," he said on another occasion. "That means to break apart. When it turns in upon itself the mind would rip itself apart." "Once I started destroying (my mind), I couldn't stop" (Sass 1992a, pp. 337–338).

All this suggests the possibility of a veritable cascade of hyperreflexivity – of the primary, consequential, and compensatory sort, and involving hyper-reflexivity of the operative as well as more reflective kind. What can result is a veritable "centrifuging" of the self – a process whereby phenomena that would normally be "inner" or tacit are progressively spun outward and away, thereby depriving the individual of the very medium of normal forms of ipseity or self-experience.

Conclusion: Phenomenological Causality

As we have just seen, phenomenological analysis of psychopathology clearly needs to integrate the three forms of diachronic explanation – by showing, e.g., how hyper-reflexivity of the primary and operative type can lead to consequential as well as compensatory forms, and how these, in turn, can exacerbate the entire process. A full analysis will, however, also need to recognize that the synchronic and diachronic dimensions laid out in this paper can and should be combined into a more encompassing, integrated or dialectical account. After all, structures are themselves formed; and in turn they play a role in developmental processes that occur over time.

A particular delusion can be understood, e.g., not *only* as an emblematic expression of a particular state of ipseity or worldhood, but also as being the causal outcome of a chain of interacting processes (thereby integrating the synchronic and diachronic perspectives). This seems to be the case with the solipsistic, influencing-machine delusion of the patient Natalija, which was mentioned above as bearing an *expressive* relationship to ipseity disturbance. A close reading of the clinical case report suggests that this patient's delusion was the outcome of a developmental process in which, over time, more primary and operative forms of ipseity disturbance led to, and inspired, forms of hyper-reflexivity (consequential as well as compensatory) that eventually crystallized in this particular delusion, with all its internal, mutually implicative, complexity (Sass 1992a, p. 227).

It is also important to consider how synchronic structures involving phenomenological implication may constitute the key context or contexts *within which* causal and diachronic processes may play themselves out. One must recognize, for example, that any alteration, over time, in degree or type of hyper-reflexivity (understood as a diachronic process) will necessarily bring along with it concomitant alterations of self-affection (because of the synchronic, equiprimordiality factor – the fact that hyperreflexivity and diminished self-affection are complementary aspects of a single whole). It will therefore be necessary also to track and describe mutations in the realm of self-affection, and to consider, as well, both the consequential and compensatory *sequelae* that *these* changes may inspire.

Another significant point concerns the mind/brain relationship. It is important to realize that the forms of symptomatic progression I have been describing cannot be considered to be *mere* epiphenomena of neurophysiological changes; indeed, they can be neither understood nor explained without making reference to the subjective or phenomenological dimension. This is not to deny the key role of neurobiological abnormalities. Indeed, these latter may well have ultimate causal primacy – as the main source of the early experiential abnormalities of the "basal irritation" (see end of section Primary Hyperreflexivity above). Once the field of experience is transformed, however, this gives rise to forms of attention and modes of experience involving developments-from or reactions-to subjectively experienced aspects of both self and world. It is not, e.g., neural events *per se* but, rather,

the *experience* of certain kinesthetic sensations as focal objects that elicits ever more intense forms of reflective concentration. In this way, subjective *experience* can play an important causal role in the progressive experiential transformations of a developing schizophrenic illness. These relevant features of subjective life concern its overall look or "feel"; they are not analyzable in terms of sentence-like propositional contents.

These experiential transformations will certainly be *accompanied* by changes on the neural plane; indeed, they may be, in large measure, manifestations on the phenomenal level of progressive organic changes occurring on the biological level. Still, these phenomenal changes are not in a purely epiphenomenal role, given that certain irreducible features of subjective life seem to provide both the motivation and the field of possibility for the progressive developments. (There may also be processes of "downward causation": when alterations on the subjective and psychological level entrain parallel changes on the neurophysiological level.) Here one might speak of a certain "autonomy of the phenomenological." As the philosopher McClamrock (1995) points out in a book on causal explanation in cognitive science, causal analysis sometimes requires one to specify a set of objects and goals that are a function of the way in which the world is *experienced* by the patient. In this sense, "irreducibly subjective" properties are sometimes able to account for a person's behavior in a way that reference to the state of the nervous system alone could not possibly do; they will sometimes constitute the "preferred level of explanation" (p. 42).[21] Husserl (1989) made the same point when he contrasted "motivational causality" with the "natural causality" of the physical world (pp. 227, 241): "The Object stimulates me in virtue of its *experienced properties* and not its physicalistic ones," wrote Husserl. "The world [that motivates my action and mental activity] is *my surrounding world*. That is to say, it is not the physicalistic world but the thematic world of my, and our, intentional life (including what is given to consciousness as extra-thematic … my thematic horizon)" (pp. 228, 230).

To clarify motivational causality is (among other things) to specify the person's (or patient's) way of seeing things and to grasp how the *perceived* environment solicits or elicits further forms of action and perception – which, of course, have their own consequences, thereby leading to comprehensible and predictable (but not wholly determined) progressions of behavior and experiential modes. It is not enough to say, then, that the experiential phenomenology of abnormal experiences merely *constrains* explanations on the cognitive or neurobiological levels: it can actually provide a key element of the explanations themselves.

---

[21] McClamrock (1995) describes "the characterizations of the world under which behavior is systematic with respect to it" (p. 4) as "distal causes", and states that, in causal analysis, these may "screen off" (i.e., render less relevant) more proximal causes (p. 54). See also pp. 45–53, 178, 187 and *passim* for arguments from philosophy and cognitive science. For related discussion, see Searle (1983), espec. pp. 112–140.

Conclusion

I have outlined a variety of ways in which phenomenology, rather than being merely descriptive, can actually have explanatory significance. Within the realm of simultaneous phenomena (the synchronic realm), I discussed *equiprimordial*, *constitutive*, and *expressive* relationships. These involve not causation but a kind of phenomenological implication. Within the realm of successive phenomena (the diachronic dimension), I considered *primary*, *consequential*, and *compensatory* processes. All three are relevant to a causal or developmental account of the genesis of schizophrenic symptoms over time. Finally I described the need to synthesize all these forms of explanation into an integrated account, and to recognize the independent role that subjectivity may play in determining forms of both experience and behavior.

References

Artaud A (1965) In: Jack H (ed) Antonin Artaud anthology. City Lights Books, San Francisco, CA

Artaud A (1976) In: Sontag S (ed) Antonin Artaud: selected writings (trans: Weaver H). Farrar, Straus, & Giroux, New York

Audi R (ed) (1999) The Cambridge dictionary of philosophy, 2nd edn. Cambridge University Press, Cambridge

Bernet R, Kern I, Marbach E (1993) An introduction to Husserlian phenomenology. Northwestern University Press, Evanston, IL

Blankenburg W (1971) Der Verlust der Natürlichen Selbstverständlichkeit: Ein Beitrag zur Psychopathologie Symptomarmer Schizophrenien. Ferdinand Enke Verlag, Stuttgart. French translation (1991): La Perte de L'Évidence Naturelle: Une Contribution à la Psychopathologie des Schizophrénies Pauci-Symptomatiques (trans: Azorin J-M, Totoyan Y). Presses Universitaires de France, Paris (Note: Both versions are cited in text above)

Bovet P, Parnas J (1993) Schizophrenic delusions: a phenomenological approach. Schizophrenia Bul 19:579–597

Buytendijk FJJ (1987) The phenomenological approach to the problem of feelings and emotions. In: Kockelmans JJ (ed) Phenomenological psychology: the Dutch school. Martinus Nijhoff, Dordrecht, pp 195–207

Carman T (2005) On the inescapability of phenomenology. In: David WS, Amie LT (eds) Phenomenology and the philosophy of mind. Clarendon, Oxford, pp 67–89

Crane T (1995) Causation. In: Grayling AC (ed) Philosophy: a guide through the subject. Oxford University Press, Oxford, pp 184–194

Cummins R (2000) How does it work? Versus what are the leaps?: two conceptions of psychological explanation. In: Keil FD, Wilson RA (eds) Explanation and cognition. MIT Press, Cambridge, MA

Davidson D (1980) Essays on actions and events. Clarendon, Oxford

Depraz N (2001) Lucidité du corps. Kluwer, Dordrecht/Holland

Dreyfus H (2002) Intelligence without representation – Merleau-Ponty's critique of mental representation. Phenomenol Cogn Sci 1:367–383

Dreyfus H, Wakefield J (1988) From depth psychology to breadth psychology: a phenomenological approach to psychopathology. In: Messer S, Sass L, Woolfolk R (eds) Hermeneutics and psychological theory. Rutgers University Press, New Brunswick NJ, pp 272–288

Dworkin RH, Oster H, Clark SC, White SR (1998) Affective expression and affective experience in schizophrenia. In: Lenzenweger MF, Dworkin RH (eds) Origins and development of schizophrenia. American Psychiatric, Washington, DC, pp 385–424

Evnine S (1991) Donald Davidson. Stanford University Press, Stanford, CA

Geertz C (1973) The interpretation of cultures. Basic Books, New York

Gray JA, Feldon J, Rawlins JN, Hemsley DR, Smith AD (1991) The neuropsychology of schizophrenia. Behav Brain Sci 14:1–20

Griffiths PE (1997) What emotions really are. University of Chicago Press, Chicago, IL

Grunbaum A (1993) Validation in the clinical theory of psychoanalysis. International Universities Press, Madison, CN

Gurwitsch A (1964) The field of consciousness. Duquesne University Press, Pittsburgh, PA

Heidegger M (1962) Being and time (trans: Macquarrie J, Robinson E). Harper & Row, New York

Heil J, Mele A (eds) (1993) Mental causation. Clarendon, Oxford

Hemsley DR (2005) The development of a cognitive model of schizophrenia: Placing it in context. Neuroscience and Biobehavioral Reviews 29:977–988

Henry M (1973) The essence of manifestation (trans: Etzkorn G). Martinus Nijhoff, The Hague. Original in1963: L'essence de la manifestation. P.U.F., Paris

Husserl E (1977) Phenomenological psychology (lectures, summer semester, 1925) (trans: Scanlon J). Martinus Nijhoff, The Hague

Husserl E (1983) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: first book (trans: Kersten F). Kluwer, Dordrecht/Holland

Husserl E (1989) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: second book: studies in the phenomenology of constitution (trans: Rojcewicz R, Schuwer A). Kluwer, Dordrecht/Holland

Husserl E (1999) In: Welton D (ed) The essential Husserl. Indiana University Press, Bloomington, IN

Jaspers K (1963) General psychopathology (trans: Hoenig J, Hamilton MW). University of Chicago Press, Chicago, IL

Kim J (1985) Psychophysical laws. In: LePore E, McGlaughin B (eds) Actions and events: perspectives on the philosophy of Donald Davidson. Oxford University Press, Blackwell

Klosterkötter J, Schultze-Lutter F, Gross G, Huber G, Steinmeyer EM (1997) Early self-experienced neuropsychological deficits and subsequent schizophrenic diseases: an 8-year average follow-up prospective study. Acta Psychiatr Scand 95:396–404

Lawson-Tancred H (1995) Ancient Greek philosophy II: Aristotle. In: Grayling AC (ed) Philosophy: a guide through the subject. Oxford University Press, Oxford, pp 398–439

Mackie JL (1974) The cement of the universe: a study of causation. Clarendon, Oxford

Marbach E (1993) Mental representation and consciousness. towards a phenomenological theory of representation and reference. Kluwer, Dordrecht

McClamrock R (1995) Existential cognition: computational minds in the world. University of Chicago Press, Chicago, IL

Merleau-Ponty M (1962) The phenomenology of perception (trans: Smith C). Routledge & Kegan Paul , New York (reissued 2002). Orig: (1945). Phénomenologie de la Perception. Gallimard, Paris

Merleau-Ponty M (1968) The visible and the invisible (trans: Lingis A). Northwestern University Press, Evanston IL

Michotte A (1963) The perception of causality. Basic Books, New York

Minkowski E (1997) Au-Delà du rationalisme morbide. Harmattan, Paris

Moran D (2000) Introduction to phenomenology. Routledge, London

Parnas J (2000) The self and intentionality in the pre-psychotic stages of schizophrenia. In: Zahavi D (ed) Exploring the self: philosophical and psychopathological perspectives on self-experience. John Benjamins, Amsterdam, pp 115–147

Parnas J (2003) Self-disorders in schizophrenia: a clinical perspective. In: Kircher T, David A (eds) The self and schizophrenia: a neuropsychological perspective. Cambridge University Press, Cambridge

Parnas J, Zahavi D (2002) The role of phenomenology in psychiatric classification and diagnosis. In: Maj M, Gaebel W, Lopez-Ibor JJ, Sartorius N (eds) Psychiatric diagnosis and classification (World Psychiatric Association Series). Wiley, New York, pp 137–162

Phillips WA, Silverstein SM (2003) Convergence of biological and psychological perspectives on cognitive coordination in schizophrenia. Behav Brain Sci 26:65–137

Polanyi M (1964) Personal knowledge. Harper Torchbooks, New York

Polanyi M (1967) The tacit dimension. Anchor Books, Garden City, New York

Ricoeur P (1992) Oneself as another (trans: Blamey K). University of Chicago Press, Chicago, IL

Sass L (1992a) Madness and modernism: insanity in the light of modern art, literature, and thought. Basic Books, New York (Harvard Paperback 1994)

Sass L (1992b) Heidegger, schizophrenia, and the ontological difference. Philos Psychol 5:109–132

Sass L (1994) The paradoxes of delusion: Wittgenstein, Schreber, and the schizophrenic mind. Cornell University Press, Ithaca, NY

Sass L (1998a) Schizophrenia, self-consciousness, and the modern mind. J Conscious Stud 5:543–565

Sass L (1998b) Ambiguity is of the essence: the relevance of hermeneutics for psychoanalysis. In: Marcus P, Rosenberg A (eds) Psychoanalytic versions of the human condition and clinical practice. New York University Press, New York, pp 257–305

Sass L (2001a) Self and world in schizophrenia: three classic approaches in phenomenological psychiatry 8:251–270

Sass L (2001b) Wittgenstein, Freud, and the nature of psychoanalytic explanation. In: Allen R, Turvey M (eds) Wittgenstein, theory, and the arts. Routledge, London, pp 253–295

Sass L (2003a) Negative symptoms, schizophrenia, and the self. International Journal of Psychology and Psychological Therapy 3:153–180

Sass L (2003b) Schizophrenia and the self: hyperreflexivity and diminished self-affection. In: Kircher T, David A (eds) The self in schizophrenia: neuropsychological perspectives. Cambridge University Press, Cambridge, pp 242–271

Sass L, Parnas J (2003) Schizophrenia, consciousness, and the self. Schizophr Bull 29:427–444

Sass L, Parnas J (2007) Explaining schizophrenia: the relevance of phenomenology. In: Chung MC, Fulford KWM, Graham G (eds) Reconceiving schizophrenia. Oxford University Press, Oxford, pp 63–95

Searle J (1983) Intentionality. Cambridge University Press, Cambridge

Simon HA (2000) Discovering explanations. In: Keil FC, Wilson RA (eds) Explanation and cognition. MIT Press, Cambridge, MA, pp 21–60

Sokolowski R (2000) Introduction to phenomenology. Cambridge University Press, Cambridge

Steinbock AJ (1995) Home and beyond. Northwestern University Press, Evanston, IL

Störring G (1987) Perplexity. In: Cutting J, Shepherd M (eds) The clinical roots of the schizophrenia concept. Cambridge University Press, Cambridge, pp 79–82 (orig. published 1939)

Tatossian A (1997) La Phénoménologie des Psychoses. L'Art du Comprendre, Paris (Juillet 1997, Numéro double, hors série)

Tausk V (1933) On the origin of the "influencing machine" in schizophrenia. Psychoanal Q 2:529–530 (orig. published 1919)

Taylor C (1993) Engaged agency and background in Heidegger. In: Guignon C (ed) The Cambridge companion to Heidegger. Cambridge University Press, Cambridge, pp 317–336

Von Wright GH (1971) Explanation and understanding. Cornell University Press, Ithaca, NY

Wakefield J, Eagle M (1997) Psychoanalysis and Wittgenstein: a reply to Richard Allen. Psychoanal Contemp Though 20:323–351

Wilson RA, Keil FC (2000) The shadows and shallows of explanation. In: Keil FC, Wilson RA (eds) Explanation and cognition. MIT Press, Cambridge, MA, pp 87–114

Wrathall MA (2005) Motives, reasons, and causes. In: Carman T, Hansen MBN (eds) The Cambridge companion to Merleau-Ponty. Cambridge University Press, Cambridge, pp 111–128

Zahavi D (1999) Self-awareness and alterity: a phenomenological investigation. Northwestern University Press, Evanston, IL

Agency with Impairments of Movement

Jonathan Cole

Beginnings

When in 1833 Sir Charles Bell (1833) first described movement and position sense (which Sherrington was to term proprioception 60 or so years later), he immediately also saw that often we do not attend to movement.

'… we use our limbs without being conscious, or at least, without any conception of the thousand parts which must conform to a single act … we stand by so fine an exercise of this power, and the muscles are, from habit, directed with so much precision and with an effort so slight, *that we do not know how he stand.* But if we attempt to walk on a narrow ledge, or stand in a situation where we are in danger of falling we become subject to apprehension; the actions of the muscles are magnified and demonstrative to the degree in which they are excited.' Bell 1833 (italics added.)

He was aware that even similar actions can sometimes be automatic and sometimes attended to. The effects of removal of these senses of movement and position sense and of touch were investigated in 1895 by Sherrington and Mott. This was possible by sectioning the dorsal, sensory, roots of a series of monkeys.

'After section … the monkey no longer used its arm or leg; movements of the hand and foot are practically abolished; the movement of grasping both with foot and hand never occurred' (Mott and Sherrington 1895).

If feeding time was deferred and the animal offered food with the intact arm tied behind its back, then it still did not use its affected limb, choosing instead to thrust its neck forward. When it tried to use its insentient foot it would miss its object widely. This profound sensory ataxia, or unsteadiness, was also seen in those pre-antibiotic days in patients with syphilitic tabes dorsalis. Mott and Sherrington then moved onto an extraordinary suggestion.

'in the case of certain movements, e.g. grasping movements of the hand and foot … the animal is rendered absolutely powerless to perform them … Although we are aware of the danger of introducing terms relating to consciousness into descriptions

J. Cole
Clinical Neurophysiology, Poole Hospital, Poole BH15 2JB, UK and Centre of Postgraduate Medical Research and Education, University of Bournemouth.

based solely on motor re-actions, we believe that we cannot more lucidly state the condition of the animals than by saying that *the volitional power for grasping with the hand had been absolutely abolished by the local loss of … sensibility*. Further this volitional power was lost immediately from the time of operation.' (italics added.)

They subsequently checked whether the deafferented cortex was still able to produce movements, by using absinthe to induce epilepsy, and found that, if anything, movements were more easily elicited. Lack of movement was not due to a cortical degeneration or loss of functioning but something before that. Interestingly, in amputees with phantom limb sensation but without the ability to move the phantoms in their mind, Sirigu's group have recently shown with transcranial magnetic stimulation that the remaining stump movements are preserved, so that the blockage in the ability to move the phantom in these patients is 'premotor' too, (Mercier et al. 2006; Reilly et al. 2006).

Sensory loss not only prevented controlled movement of the limb, but also abolished the will or at least the conscious motor focus or intention to move it. Such observations, made on monkeys at the beginning of the twentieth century, had to wait for clinical studies on those with severe sensory loss to be confirmed or refuted nearly a century later. But in their statement Mott and Sherrington, two empirical physiologists have also shown that observations about agency and action can be made from situations when the normal implicit relations between these break down.

When we move we usually focus on high level goals rather than the moving part or even the action. We make a cup of tea and drink it, we don't think how we make it or how we move the arm and hands to do so. The body is often transparent or absent from our attention, (Leder 1990; Gallagher 1986). The actions we make might be divided into several broad groups, instrumental ones; dressing, writing etc, locomotion; walking, running or even driving, autonomic or body maintenance ones; washing, defaecating, etc., and communicative action; speech, gesture and facial expression. This account will give examples of how difficulties in action due to neurological impairment and paralysis affect the perception of agency and embodiment and how these various difficulties affect these types of action differently. It will also be clear that agency and independence of action is always placed in a wider context, of self and others. Movement by its nature is embodied, but this embodied action is how we define ourselves to a significant extent, so that difficulties in agency are not only perceived, but also affect the agent's self-esteem.

But before this we will review, briefly, some empirical recent work on the perception of action and agency.

Empirical Observations on Will and Action Awareness

Wittgenstein made many observations on the relation between action and will, and between voluntary and involuntary movements.

> Writing is certainly a voluntary movement, and yet an automatic one. And of course there is no question of a feeling of each movement in writing. One feels something, but could not possibly analyse the feeling. One's hand writes; it does not write because one wills, but one wills what it writes.

> One does not watch it in amazement or with interest while writing; does not think "what will it write now?" but not because one had a wish it should write *that* for that it writes what I want might very well throw me into astonishment (Wittgenstein 1981 p. 586).

> 'How could I prove to myself that I can move my arm voluntarily? Say by telling myself: "now I will move it" after which it moves? Or should I say "simply by moving it"? How do I know that I did it, and that it did not move just by accident? Do I in the end feel it after all?' (Wittgenstein 1980, p. 595).

The relation between intention and movement were investigated by Libet et al. (1983) in experiments which are still controversial in their interpretation. They asked subjects to watch a clock rotating every 2.5 s and make a finger movement voluntarily, when they chose, and report where the clock was when they became aware of the will to move, (the W judgement, the moment of conscious intention). In other experiments they were asked to judge when they actually moved, the M judgement. The timing of the action was taken from EMG activity from the relevant muscle. Lastly they also recorded motor preparatory potentials, the readiness potential, from the brain prior to the action.

The readiness potential was seen 1,000–500 ms before the movement, the W judgment 206 ms before movement and the M moment 86 ms before movement. Some have questioned the nature of free will and voluntary action because of such results, since intention seems to follow motor preparation and not precede it. More recently these results have been replicated and extended by Haggard and Eimer (1999), who showed that conscious intention is related to the lateralised readiness potential which is related to a specific movement, e.g. move the right hand. Haggard suggests that awareness may have evolved to be tied to a process related to specific movement selection, (see Haggard and Libet 2001).

Wittgenstein continued,

> "Willing, if it is not to be a sort of wishing, must be the action itself. It cannot be allowed to stop anywhere short of the action." If it is the action, then it is so in the ordinary sense of the word; so it is speaking, writing, etc. But it is also trying, attempting, making an effort, – to speak, to imagine, etc. (Wittgenstein 1980, p. 615)

Recently the study of perception has broadened from analysis of sensory phenomena to include the analysis of the perception of action. This work has reached some broad conclusions; when we move much of what we perceive is tagged to our intention to move rather than to our perception of what has happened as a result of movement, at least for sensation predicted from that action, (Fourneret and Jeannerod 1998). By measuring the timing of perception of an intentional action and of its sensory consequences, and by disturbing these with transcranial magnetic stimulation, Haggard and colleagues showed how these two percepts are bound together (Haggard and Magno 1999; Haggard et al. 2002; Haggard and Clark 2003). Conscious experience of action may be related in varying degrees to the process of action execution in a predictive manner as well as a reconstruction of intentional action based on retrospective inference.

Awareness of voluntary action or a sense of agency, that sense of intending and executing an action, appears to derive from a stage later than intention but earlier than movement itself, (see Marcel 2002), at least for short duration intentional

action in an experimental situation. One might distinguish between this sense of agency and a sense of ownership, that it is one's own body which experiences a sensation, (see Gallagher 2005). Recent work has begun to untangle the relations between these two and show how intentional action may modulate incoming perception (see Tsakiris and Haggard 2005).

These excellent studies depend on acute experiments in control subjects; in daily life intentions towards movement and action proceed with varying amounts of attention to them at various stages of their unfolding. Marcel also considered the longer term, perhaps more physiological or ecological, sense of agency as background, and suggested that action contains a temporal dimension beyond itself with a sense of being an agent having past, present and future aspects.

'The normal sense of ownership is due to our non-observational awareness of action; pathological ownership of action is due to the person's awareness of being observational' (Marcel 2002).

Sometimes the relation between agency and action may break down either due cortical disconnection syndromes, like alien and anarchic hand, (see Farrer and Frith 2002; Gallagher 2004) or due to inaction and paralysis due to breakdown of the sensory and motor apparatus itself. It is to these latter that we now turn. How is Marcel's awareness of being observational made manifest in impairment of movement?

Immediate Perceptions of Paralysis and Inaction

The most and acute complete paralysis imaginable is spinal cord injury; when people break their neck they know immediately what they have done. Their thoughts are often not about action or agency but about self and future. After such an accident one person told of her first thoughts,

> I had completely ruined my life … I was going to be a burden to everyone … My husband would probably leave me – or, if he stayed, it would only be out of some noble sense of obligation. The children would grow up with the social impediment of a wheel chair bound mother. I was a useless cripple. Disabled, damaged goods. (Hill 2000, p. 34)

Another said,

> The first thing I assessed was what I could no longer do. What the future did not hold for me. You are totally self-obsessed at the beginning. I will not raise a family; I will not be independent. Someone else will have to do my bowel care, put me to bed and wash me (Cole 2004a, p. 81).

These thoughts were about the loss of independent movement but also about themselves as agents, leading some to question their own continued validity as individuals. Yet others, immediately, begin to explore their new bodies and their new lives. One man told me,

> I was concerned with the practicalities, I suppose, of getting on with life. I didn't lie there thinking all the time "Oh My God what have I done, what's this going to mean?" I never burst into tears because, from the early stages of living with the injury, I have seen the

whole thing as a challenge. How do I overcome so and so? How do I deal with this? I never thought, "I can't do that". (Cole 2004a, p. 66)

These people seem to have such a strong sense of themselves as agent, and as self, together with an openness to new experience that can withstand even paralysis.

Such an enormous loss in sensory input as spinal cord injury might be expected to have some lower level sensory effects and this is, indeed, the case. Immediately after the injury people often feel their limbs to be in unusual positions, though they cannot feel them. One woman, twenty minutes or so after her injury, was distressed when the emergency services arrived because she thought her legs were sticking up in front of her, revealing more than she wanted. *'Please put my legs down, they're in the air'*. This is not uncommon and seems to be an immediate neurological response to sensory loss.

Later some with loss of sensation from below the neck feel disconnected and disembodied. One tetraplegic related,

> Early on there was this sensation of feeling nothing and one felt disembodied, even though you knew you weren't because you could see. But you did feel completely disembodied. (Cole 2004a, p. 31)

Another suggested, curiously, that his pain, felt to be in the body he could not feel (a form of phantom body pain), gave him an illusion of connection,

> My physical pain is in the hands and down the legs and in the feet. The pain does not come on; it is there, the whole time, 24 h per day every day, every day of the year. Most of the time I can override it, but I still know it is there. But when it's angry it gets me so down. I cannot do anything myself to get a connection [with my body]. If I pinch my legs it is numb. [When the pain is easier] it is almost comfortable, almost my friend. I know its there, it puts me in touch with my body. The pain is the connection – my friend the pain. (Cole 2004a, p. 89)

Most people view their insentient and paralysed body as being theirs. One man was discussing his pain, in a similar way to the previous man, when he went on,

> I would not want to be without some buzzing [mild pain]. It gives me a sense of identity in my body. I still view my body as whole; it's just motionless. I'm not a head on a bag of potatoes. I still know it's there, I still like it; I like to see it, it is still me, and I am still it, totally. I always say at the end of the day I am motionless, but I am still me, body and all. (Cole 2004a, p. 24)

Agency After Pure Sensory Loss

Some of the relations between agency, intention and movement are revealed from accounts of those with the rare acute sensory neuronopathy syndrome. In this the nerves underpinning cutaneous light touch sensation and proprioception are destroyed, over a few days, leaving the person in a situation like Sherrington's monkeys, with the potential to move, since movement or motor nerves are unaffected, but with no feedback of movement from the limb (though perception of pain

and temperature persist). This initially renders all movements inaccurate and ataxic, but also affects, of course, the perception of action.

Ian Waterman, (IW), had such a neuronopathy aged 19 and has had no recovery, over 30 years later. Initially bed bound and unable to control movement, or to feel his body, he felt disembodied, as had Sacks's case (Sacks 1981). This feeling retreated once he had discovered a way of moving using cognitive effort and visual control. Once he had learnt to control movements, his sense of being in his body and what might be broadly called embodiment he describes as normal, i.e. similar to before. In contrast GL, another person with the neuronopathy, has a level of sensory loss that is higher, from the lower face, and has not been able to impose her will on movement to the same degree (for many reasons). She relates that her relationship to her body is more like that of a captain to a boat, controlling but at a distance. Their feeling of disembodiment may relate to a lack of controlled movement rather than simply, or necessarily, to the sensory loss. Once he had controlled movement, albeit cognitively, IW felt less disembodied.

Without peripheral feedback IW has to think about movements and monitor them visually the whole time to maintain accuracy. Few movements are automatic skills and all require some cognitive control. His focus is not only the goal but the means, the need and the action. He first realised that this approach would be needed 6 weeks after his deafferentation. Lying in bed he thought that if he contracted his stomach and looked at it then he would sit up. When this happened he was so pleased that he no longer concentrated and fell back uncontrollably. The moment was double edged; the realisation that movements were possible again was wonderful but he also knew they would need mental concentration on them to a huge degree.

Over the next 17 months, as a patient in a rehabilitation ward, he learnt to dress, feed himself and eventually, after a year, to stand and walk. All the while he was looking and thinking about action. He spent 6 weeks making a jigsaw to teach himself how to use the fingers again, also making paper clip chains endlessly for the same reason. He would eat food which had gone cold rather then be fed. On the one hand movements he had made before the illness were easier after, but on the other he was very aware that something he had managed one day was not necessarily possible the next. Once having got drunk he found himself unable to move for days, so fragile was his attention to movement. If he sneezed he would fall over, day-dreaming whilst standing was impossible (Cole 1995).

Slowly, over years, some movements have become marginally easier. He now says that to walk, in a well lit, un-crowded, level, non-slip, wind free space, takes round 50% of his mental concentration and gesture appears at times to be automatic. But, equally, if he has a head cold he cannot think well enough to function and so goes to bed. Recently, over 30 years after the illness, after sleepless nights with an infection, he could not coordinate well enough to reach for a glass, revealing how vulnerable his ability to movement was.

Once when sitting at a table, he described his level of attention when reaching for a cup of tea,

> I am initially being aware of my body position to hang it all off. Sitting down my legs are in a tripod, sitting on a chair, and I have a mental image of this. Having my arm resting on the table is a good triangle position, and I know I can reach the cup. Once the framework is safe and that I can monitor hand out and in then I can begin. I can see it all except what the fingers are doing behind the cup, but I have learnt to do this [by grasping without using the handle]. I don't know how heavy the cup is but know roughly, pick it up and monitor it visually. I need to see my arm up to my face, but then I can feel the cup with my face. I may not need to see the cup return until it reaches the table. This is all very controlled and involved.

> In the early days every finger was monitored. Now am I safe with hand movements. When I move the hand it is almost automatic that the fingers open though I keep the lateral three fingers out of way. To make any reaching movement though first I tighten up the back muscles to make the start safe, and then move the arm out, less hand open but watch it to make sure it is doing what you want (Cole and Waterman, unpublished observations).

Before making any movement he rehearses it in his mind, his 'pre-visualisation.' All this means that IW has to really want the tea to make the movement worthwhile. The effort involved in these movements and his insistence on doing things at the limit of his mental concentration is very tiring. Being tired is measured by his ability to think sharply enough about movement. When in a restaurant I asked how he would get up to leave.

> I cannot concentrate on all aspects of walking. In a given movement sequence, say from here to the door in an empty restaurant over a distance of 20 ft, I think two steps ahead roughly. If it is busy I will sit and wait. Walking is never the same each time. I apply the strategy which is easiest at the time. These are not always major changes, but like a snooker player, some days I have flair and some days not.(Cole and Waterman, unpublished observations).

He drives, using hand controls, and finds it easier to do this for several hours than get out to fill up with petrol.

> Inner achievements are all; goal setting, pre-visualisation. I cannot be too intellectual about it though, because all my intellectual effort is in doing it not reflecting on it. I can't spend too long thinking about it, because if I thought what I have to do all my life I would be too scared. (Cole and Waterman, unpublished observations)

For various reasons, level of deafferentation, age, ability to produce motor programmes, (Miall and Cole 2007), IW's recovery of movement is unusual for those with deafferentation, and most of the few with neuronopathy and such severe sensory loss due to other causes live from wheelchairs and remain quite ataxic in their movements.

Forgetting How To

Initially IW was not only unable to control movements without feedback; he also had profound problems in knowing what he had to do to move. He knew intellectually, of course, what movements were needed, but found it difficult to form ideas, and so motor commands, in his head about movement. His ideational programs for action were no longer accessible to him, or were themselves degraded without

feedback. Without peripheral feedback the link between thought and action had been disrupted. If his motor programmes were disabled without feedback, then his conscious focus onto and choosing of motor commands were also affected, giving him a sort of ideational apraxia due to deafferentation.

There are other examples of this. A colleague who fractured his heel was, once out of the plaster and pain free, surprised to find that he could not move his foot. He felt he had 'forgotten what to do.' For several months automatic foot movements were lost. Another subject, suffering from viral meningitis, tried to move his neck, which turned out to be stiff though not painful. When he went to move the neck not only did it not move; he too felt that he no longer knew how to move it (Cole 2004b). A similar phenomenon was described by Sacks after breaking his quadriceps tendon in the thigh and being unable to move his leg for some months (Sacks 1984).

Such a phenomenon has also been described after central motor damage. Brodal, a neuroanatomist who suffered a purely motor stroke, was initially paralysed down one side. When he had re-learnt some movements later he discussed tying his bow tie, something he had done previously every day for 40 years, thus,

> The finger movements were difficult to perform with sufficient strength, speed and co-ordination, but it was quite obvious that the main reason was something else. Normally the necessary numerous small delicate movements follow each other in proper sequence almost automatically, and the act proceeds without attention. Subjectively I felt as if I had to stop because the fingers did not know the next move. (Brodal 1973)

Withdrawal of sensory feedback, stiffness in a joint, pain or disuse, or motor cortex damage, may present itself to consciousness as a difficulty in the focussing of attention into action commands. Action or, more correctly, the merging of thought – conscious intention – into action seems to be dependent on the state of the peripheral sensory apparatus and feedback from it, and on the motor cortex, in ways similar to some with apraxia due to cortical damage. Problems are presented subjectively as not just inability to move but reductions in the possibilities of commanding or imagining those movements. When IW began to learn how to move he would rehearse and imagine these movements in his mind before doing them, flipping from not knowing how to, to a condition where this 'knowing how to' preceded the action and improved it.

Agency in Paralysis

There may be few larger disruptions of agency and intention than complete, permanent, paralysis. Robert Murphy, a professor of anthropology, became quadriplegic late in life (Murphy 1987). His paralysis had profound effects at social, ontological levels but he also revealed how it affected his own intentional motor system.

> for a while I tried to will the legs to move, but each futile attempt was psychologically devastating … I was saved from the edge of breakdown because the slow process of paralysis of my limbs was paralleled by a progressive atrophy of the need and impulse for physical activity. I was losing the will to move. (Murphy 1987, p. 78)

Later after quadriplegia illusions of action may also occur. Christopher Reeve wrote of how,

> As I started to face reality during the month in intensive care and six months in rehab, moments from my former life kept popping into my head. It was like a slide show, but the pictures were all out of sequence. As a long plastic tube was inserted through my neck and pushed down into my lungs to remove accumulating fluid, suddenly I would be sailing in Maine. A moment or two later, Dana and I were making love; I was on a horse jumping over stone walls in the countryside; I was taking a curtain call after a performance in the theatre; carrying boxes and lugging furniture up four flights of stairs into my first apartment in New York …

> [Later] … more images flashed onto the screen, usually snapshots of my most cherished memories when I was whole and healthy and free. (Reeve 2002, p. 12).

These may be more like the delayed higher ideations that have been described after acute visual loss in Charles Bonnet syndrome (ffytche and Howard 1999). In some people the movement brain may not fall quiescent once deprived of action. Reeve's account did not go further into the phenomenology of how owned these actions were, nor how immersed Reeve was within them (though one hopes he was).

In another, filmed, study, we interviewed a man, Michael, paralysed from the neck down with ALS. A cello teacher for many years with the paralysis he was unable to bear to listen to music for about a year. But then he found ways of enjoying it once more. He found when listening that the sheet music appeared as though just above him in mid air, visually, and that he had the similar immersed hallucination that he felt he was playing it too, even though paralysed. The movement brain and the music brain were combining to give him, at one level, an immersed illusion of agency once more (Cole et al. 2007).

Murphy's later reflections show a retreat from the world and from attempts to act and interact with it,

> a quadriplegic's body can no longer speak a 'silent language' … the thinking activity can no longer be dissolved into motion, and the mind can no longer be lost in an internal dialogue with physical movement … My thoughts and sense of being alive have been driven back into my brain … many say they are no longer attached to their bodies … my former sense of embodiment remained taken for granted … my sense of re-embodiment is problematic negative and conscious … consuming consciousness of handicap even invades one's dreams. Even in sleep disability keeps its tyrannical hold … (Murphy, op cit, p. 87)

> The totality of the impact of serious physical impairment on conscious thought … gives disability a far stronger purchase on ones sense of who and what he is than do any social role … which can be manipulated. Each social role can be adjusted to the audience, each role played before a separate audience, allowing us to lead multiple lives. One cannot however shelve a disability or hide it … It is not a role: it is an identity … society will not let him forget it. (Murphy, op. cit. p. 90)

Moving though this is, again – fortunately – Murphy's experience does not appear typical. Many, most, people with quadriplegia do well and use whatever movements are available to them to translate intention into acts.

It is important also to realise that a tetraplegic is usually not without movement completely and that their attention to the movements they make is often more than

in control subjects. To sit in a wheelchair is not passive, but a complex and continuing balancing act; without postural muscles, and reliant on weak shoulder and proximal arm muscles, being upright in a chair can be active and tiring. Tetraplegics also have to attend to bladder and bowel care and avoid pressure sores. Though instrumental and locomotor actions may be difficult and severely reduced, their relation with their paralysed, insentient body is one of vigilance towards, and action in relation to, what was previously automatic and autonomic.

Effort After the Loss of Automatic Action

IW's mental concentration on movement has to be relentless which, in turn, is immensely tiring. He will find a place to flop, a secure chair or bed, and have 'down time' just to relax several times a day. He has described his new life as a daily marathon; a marathon of mental rather than physical effort. Brodal reflected similarly on his profound weakness after his stroke.

> It was a striking and repeatedly made observation that the force needed to make a severely paretic muscle contract is considerable. Force in this connexion refers to what might call a force of innervation. Subjectively this is experienced as a kind of mental force, a power of will. In a muscle just capable of being actively moved the mental effort needed was very great. Subjectively it felt as if the muscle was unwilling to contract, and this could be overcome by very strong voluntary innervation.

> This force of innervation is obviously some kind of mental energy, but the result is seen as contraction of the muscles in question. (Brodal 1973)

In acute experiments in which controls were paralysed whilst awake, their attempts to move led to strong sensations of effort and illusions of actual movement, not felt during willed attempts to move with the arm under peripheral ischaemic paralysis and anaesthesia. Gandevia et al concluded that the illusions of movement when paralysed might have arisen from muscle spindle receptors within muscles known to be more resistant to paralysis with the drug used (Gandevia et al., 1993). Patients however describe a more central effort related to attention to movement. Recently evidence has begun to emerge of the involvement of the anterior cingulate cortex in the subjective feeling of effort associated with action (Naccache et al. 2005, Mulert et al. 2005). One presumes that such feeling is equivalent to the qualia of sensation, tagging intentional motor commands with a central alerting percept.

Agency with Altered or Reduced Embodiment

There are some situations in which a person may have agency and intention without a moving part of the body (see Cole, 2007). One tetraplegic told me that he helps his wife do the dishes in the evening. By this he meant that his personal assistant

would do them. But by doing so in a manner and at a time controlled by the tetra-plegic, the PA allowed this tetraplegic to say he was doing it. This seemed more than a figure of speech. Since the lack of ability to act on the world is one of the major problems of those with spinal cord injury, the facilitation of action through a PA is not trivial.

Bell, amongst others, described how people after amputation may sometimes be able to move their phantom limbs, or have them move apparently without intention during gesture. Recently, following the demonstration of a mirror box by Ramachandran et al. (1995), those with phantom limbs, have been able to entrain movements in the phantom when they see the opposite existing limb mirror reversed and moved; motor intention to one limb has been conflated to the other. In an extension to this several workers have worked with virtual environments which can present a virtual limb to a subject and which can be moved either from move-ments of their existing limb, or from movements of the stump, picked up via motion sensors either on the existing limb or from the remaining stump (Desmond et al. 2006; Murray et al. 2007; Cole et al. 2009).

Within minutes of being in such a system many people have learnt to move their phantoms appropriately and not only see but also feel such movement. When they see their virtual limbs touch something in the virtual world they sometimes also feel touch. The necessary conditions for the return of agency may be visual feedback of a moving limb to entrain intention to, and with this may come a return of somatic sensation. This, though, also appears to require continuing and tiring effort towards the movement.

Affect and Inaction

Thus far agency and resulting sensation has been considered in relation to move-ment control within a cognitive framework. But, right at the beginning, Bell saw it was more,

> The exercise of the muscular frame is the source of some of our chief enjoyments. This activity is followed by weariness and a desire for rest; and although unattended with any describable pleasure of local sensation, there is diffused through every part of the frame a feeling almost voluptuous. (Bell, op. cit.)

Subjects with the acute sensory neuronopathy syndrome have intact or near intact nerves underpinning the sensations of temperature and pain, through small myelinated and unmyelinated sensory nerves unaffected by the acute sensory neu-ronopathy. Recently experiments have provided evidence that some of these small unmyelinated CT fibres are also involved in the signalling of low threshold pleasant touch. One subject, GL, when stroked gently on the forearm had little sensation of touch *per se* but did find the sensation pleasant. With this fMRI activations were seen not in the usual area for touch sensation, the sensory cortex, but in the insula cortex, an area thought to be involved in the elaboration of the somatic self, (Olausson et al. 2002). Such experiments can only be performed in subjects with a

neuronopathy since light touch in control subjects would activate large numbers of large myelinated fibres and mask any other effects. If the pleasantness of the perception of a caress may be conveyed in part via CT afferents, it seems likely that, as Bell suggested, proprioception has an affective emotional aspect too (Cole and Montero 2006).

Some of the consequences and subjective experience of the lack of intention and action have been described at levels of action awareness. Wegner (2002) and Haggard (2004), independently, have suggested that a close binding between the awareness of intention and of action helps calibrate our sense of self – what we intend to move, and do move, is us. This cognitive view of self and non-self perhaps needs another dimension. Possibly one of the most important purposes of affective proprioception is that it binds our agent or acting self with embodied self at a more emotional level.

The Communicative and Emotional Self in Action and Inaction

The goal of instrumental or locomotor actions may be relatively clear, but there is another form of embodied action, emotional and communicative action, which is rarely attended to. We gesture and make facial expressions which have profound effects on interpersonal relations with little or no awareness of what we do. The loss of these movements reveals their import.

IW and GL, having learnt instrumental actions, feeding, dressing, etc and loco-motion, either in a chair or walking, both realised that they no longer used gesture. Without this they felt isolated and disabled. So they both learnt to make gestures with speech consciously, to give the illusion of normality. Now, after 3 decades of moving without sensation, gestures are probably the most automatic and normal movements that IW makes. This may be because in part gestures are part of a thought and language system which was not affected by the neuronopathy and partly because gesture has to be accurate in shape or form and in time but not in place, and movements in accurate relation to the external world are the most diffi-cult without peripheral sensory feedback. (Cole et al. 2002).

In the rare congenital condition, Möbius Syndrome, the cardinal features are an absence of the cranial nerves responsible for lateral movement of the eyes and for movement of the muscles of facial expression. People with Möbius, from birth, are unable to move their faces or their eyes in their heads, leading to complex problems with feeding, speech and gaze. Without facial expression from birth, some people with the condition have profound problems with interpersonal communication. They also, perhaps not coincidentally, often have reduced use of gesture. One man, in his fifties, described it thus,

> I have a notion which has stayed with me over much of my life – that it is possible to live in your head, entirely in my head. Whether that came out of my facial problem I don't know. It is only very recently that the whole area of non-verbal communication has even

come to my attention. I know now that since I put out a reduced range of signals I receive back a similarly reduced range.

When he met his wife to be, he told me that,

I think initially I was thinking I was in love with her. It was some time later when I realised that I really felt in love.

I think I get trapped in my mind or my head. I sort of think happy or I think sad, not really saying or recognising actually feeling happy or feeling sad. Perhaps I have had a difficulty in recognising that which I'm putting a name to is not a thought at all but it is a feeling, maybe I have to intellectualise mood. I have to say this thought is a happy thought and therefore I am happy.

I think also that I have a fear of being out of control with emotions, feeling something that I can't manage. I have also found it very difficult to communicate feelings throughout my life, whether as a child or with my wife, though I think I am getting better at it now. I don't really know how I communicate happiness or sadness. That's a very hard question. Some people cry when they're sad. I don't. I sometimes felt that I would like to be able to cry but you see I am not really able to cry, my tears can come but there's nothing else. My tears only flow when I eat. I am afraid of such feelings. I try and shut them off. 'Of course, since I have never been able to move the face, I've never associated movement of the face with feeling of an emotion.' (Cole 1998, p. 122ff.)

Discussing her young son, Duncan, who lives with Möbius, his mother said,

By eight months people were whispering that something was wrong. My other children had been smiling and crawling around by then but Duncan didn't do anything. The first time I really understood what was going on inside his head was one day when I was looking for a nappy pin for my youngest and he said, "Me get it", and he went and fetched it from the draw. He was three then. Up until then no-one had known what intelligence Duncan might have. (Cole 1998, p. 139)

He had been completely passive, seeing and hearing, and yet unable to impose a sound or thought or emotion upon those around him. In place of the spontaneous and sometimes volcanic emotions children experience and display, and then learn to control, Duncan appeared calm and unemotional. The highlights of a normal childhood seemed to pass him by,

I remember his fifth birthday party he was sat in his high chair and went to sleep; it was just like another day for him. He didn't want to know, he didn't want to play. He doesn't really get excited on birthdays, even his own. It is difficult to know when he's having fun. When he comes home from school we don't know how he's feeling, we have to ask him. Everything is questions and answers. He has always been a very placid child. He never really gets angry, never really appears upset. (Cole 1998, p. 140)

Though this experience of paucity of emotional experience may not be seen in all those with Möbius, and is not always recognised for what it is, its presence in even a few suggests that embodied facial expression of emotional states, and social sharing of these emotions, may at least in part be necessary for their being experienced. Though facial expressions are rarely consciously intended, the relation between inaction on the face and emotional experience seems intimate and affects our development of self and of social competence. We calibrate this from how others respond, 'I exist in the facial expression of the other …' (Merleau-Ponty 1964).

Extraordinarily, some with Möbius come to recognise their emotional impoverishment and learn to experience feelings, by becoming more social, by aping others, and by developing gesture, prosody and language – anything they have – which allows them to reach out to and share conversations and feelings with others. (Cole with Spalding 2008)

Conclusions

Recent experiments point towards that the awareness of action, or sense of agency, may follow intention but precede the actual movement and show that this intentional action may also affect incoming sensation from movement. These serve to unify experience of action when its components are slightly dispersed in time. It may also allow a calibration of self, since what we move and see moving we are.

Attention towards movement is usually directed towards its higher goal. But when there is impairment of action then attention shifts to the means of movement as well as its aim. Thus IW, without sensory feedback, has to observe his moving part as well as think about aspects of movement usually made automatically. But disruption of the usual link between thought and movement, whether due to sensory loss or a motor weakness due to stroke, can also be presented to consciousness in other ways, for example in the perception of 'forgetting' what movement to make. The focussing of intention towards formulation of action at the higher level may be dependent on the integrity of the motor brain and sensory apparatus in some way too.

One consequence of moving in the face of some neurological impairments is a mental tiredness. Whether focussing on movement because sensory feedback is absent, or because of a motor stroke, the process is enormously tiring mentally, just as the learning of new motor skills can also be tiring.

If such observations are made in the context of action and movement, disruptions of these have far wider effects. The relation between thought and action were considered by the newly paralysed Murphy, who described how his thoughts and mind could no longer be lost in movement. The impact of serious physical impairment on conscious thought remains largely to be explored empirically.

The feeling of tiredness which those with impairment describe seems different to the pleasant feeling of fatigue which can follow and exist during physical exercise. The latter, however, makes us aware that action and movement are not solely perceived in cognitive motor terms but also in affective, emotional ways too. Sherrington wrote that,

> Mind rarely, probably never, perceives any object with absolute indifference, that is without 'feeling.' All are linked closely with emotion. (Sherrington 1900, p 974)

The experience of reduced agency and action affect the perception of the embodied and social self as well; problems with movement are rarely perceived as only motor problems. Deficits in agency affect the agent and at times lead people in these positions to question and doubt themselves in ways Murphy has described.

And, in turn, it is through the body that we communicate with others, through combinations of language, gesture, posture and facial expression. Some of the consequences of disruptions in these were seen in the need for relearning of gesture after sensory loss and in those paralysed by spinal cord injury. But perhaps the clearest effects are seen in those with impairments of facial expression. These show the importance of embodied non verbal communication for social existence and, indeed, for emotional experience itself. Impairments in movement may reveal not only something about our normal sense of agency but the emotional aspect of embodiment and action, and also how embodied, affective and social is our sense of self.

References

Bell, Sir C (1833) The hand: its mechanism and vital endowments as evincing design. Bridgewater Lecture Series. London. Pickering reprinted by Pilgrim's Press, Brentwood, 1979.

Brodal A (1973) Self observations and neuro-anatomical considerations after a stroke. Brain 96:675–694

Cole J (1995) Pride and a daily marathon. MIT Press, Cambridge, MA

Cole J (1998) About face. MIT Press, Cambridge, MA

Cole JD (2004a) Still lives. MIT Press, Cambridge, MA

Cole J (2004b) On the relation between sensory input and action. J Motor Behav 36(3):243–244

Cole J (2007) The phenomenology of agency and intention in the face of paralysis and insentience. Phenomenol Cogn Sci 6:309–325

Cole JD, Montero B (2006) Affective proprioception. Janus Head 9(2):299–320

Cole J, Spalding H (2008) The invisible smile. Oxford University Press, Oxford/New York

Cole J, Gallagher S, McNeill D (2002) Gesture following deafferentation: a phenomenologically informed experimental study. Phenomenol cogn Sci 1:49–67

Cole J, Dawson A, Rawlence C, Walker L (2007) Still life in eight movements. Wellcome Trust Sciart project

Cole J, Crowle S, Austwick G and Henderson-Slater D (2009) Exploratory findings in virtual induced agency for phantom limb pain. Disability & Rehabilitation, 10:846–854

Desmond DM, O'Neill K, De Paor H, Mac Derby G, MacLachlan M (2006) Augmenting the reality of phantom limbs; three case studies using an augmented mirror box procedure. J Prosthet Orthot 18(3):74–79

Farrer C and Frith CD (2002) Experiencing oneself vs another person as being the cause of an action. Neuroimage, 15:596–603.

ffytche DH, Howard RJ (1999) The perceptual consequences of visual loss; positive pathologies of vision. Brain 122:1247–1260

Fourneret P, Jeannerod M (1998) Limited conscious monitoring of motor performance in normal subjects. Neuropsychologia 36(11):1133–1140

Gallagher S (1986) Lived body and environment. Res Phenomenol 16:139–170

Gallagher S (2004) Neurocognitive models of schizophrenia: a phenomenological critique. Psychopathology 37:8–19 (with a response from Christopher F (2004) Comments on Shaun Gallagher. Psychopathology 37:20–22

Gallagher, S (2005) How the body shapes the mind. Oxford: OUP

Gandevia SC, Killian K, McKenzie DK, Crawford M, Allen GM, Gorman RB, Hales JP (1993) Respiratory sensations, cardiovascular control, kinaesthesia and transcranial stimulation during paralysis in humans. J Physiol 470:85–107

Haggard P, Clark S (2003) Intentional action: conscious experience and neural prediction. Conscious Cogn 12:695–707

Haggard P, Clark S, Kalogeras J (2002) Voluntary action and conscious awareness. Nat Neurosci 5(4):382–385

Haggard P and Eimer M. (1999) On the relation between brain potentials and the awareness of voluntary movements. Exp Brain Res. 126(1):128–33.

Haggard P and Magno E. (1999) Localising awareness of action with transcranial magnetic stimulation. Exp Brain Res. 127(1):102–7.

Haggard P and Libet B. (2001) Conscious intention and brain activity. Journal of Consciousness Studies, Volume 8, Number 11, pp. 47–64(18)

Haggard P. Conscious intention and motor cognition. Trends in Cog Science, 9, 6, 290–295

Hill J (2000) Footprints in the snow. Macmillan, London

Leder D (1990) The absent body. University of Chicago Press, Chicago, IL

Libet B, Gleason CA, Wright EW, Pearl DK (1983) Time of conscious intention to act in relation to onset of cerebral activity (readiness potential). The unconscious initiation of a free voluntary act. Brain 106:623–642

Marcel A (2002) The sense of agency: awareness and ownership of action. Chapter 2, In: Roessler J, Eilan N (eds) Agency and self-awareness. Oxford University Press, Oxford, pp 48–93

Mercier C, Reilly KT, Vargas CD, Aballea A, Sirigu A (2006) Mapping phantom movement representations in the motor cortex of amputees. Brain 129(8):2202–2210

Merleau-Ponty M. The Primacy of Perception. Illinois: Northwestern University Press

Miall RC, Cole JD (2007) Evidence for stronger visuo-motor than visuo-proprioceptive conflict during mirror drawing performed by a deafferented subject and control subjects. Exp Brain Res 176(3):432–439

Mott FW, Sherrington CS (1895) Experiments upon the influence of sensory nerves upon movement and nutrition of the limbs. Proc Roy Soc 57:481–488

Mulert C, Menzinger E, Leicht G, Pogarell O, Hegerl U (2005) Evidence for a close relationship between conscious effort and anterior cingulate cortex activity. Int J Psychophysiol 56(1):65–80

Murphy R (1987) The body silent. Henry Holt, New York

Murray CD, Pettifer S, Howard T, Patchick EL, Caillette F, Kulkarni J, Bamford C (2007) The treatment of phantom limb pain using immersive virtual reality: three case studies. Disabil Rehabil 29(18):1465–1469

Naccache L, Dehaene S, Cohen L, Habert MO, Guichart-Gomez E, Galanaud D, Willer JC (2005) Effortless control: executive attention and conscious feeling of mental effort are dissociable. Neuropsychologia. 43(9):1318–28

Olausson H, Lamarre Y, Backlund H, Morin C, Wallin BG, Starck G, Ekholm S, Strigo I, Worsley K, Vallbo AB, Bushnell MC (2002) Unmyelinated tactile afferents signal touch and project to insular cortex. Nat Neurosci 5(9):900–904

Ramachandran VS, Rogers-Ramachandran D, Cobb S (1995) Touching the phantom limb. Nature 377:489–490

Reeve C (2002) Nothing is impossible; reflections on a new life. Century, London

Reilly KT, Mercier C, Schieber MH, Sirigu A (2006) Persistent hand motor commands in the amputees' brain. Brain 129(8):2211–2223

Sacks O (1981) The man who mistook his wife for a hat. Duckworth, London

Sacks O (1984) A leg to stand on. Duckworth, London

Sherrington C (1900) Cutaneous sensation. In: Schafer EA (ed) Textbook of physiology, Pentland, Edinburgh. p 974

Tsakiris M, and Haggard P (2005) Experimenting with the acting self. Cognitive Neuropsychology, 22(3/4):387–407

Wegner D (2002) The illusion of conscious will. MIT Press, Cambridge, MA/London

Wittgenstein L (1980) Remarks on the philosophy of psychology. Blackwell, Oxford, p31e

Wittgenstein L (1981) Zettel, 2nd edn. Blackwell, Oxford

Index