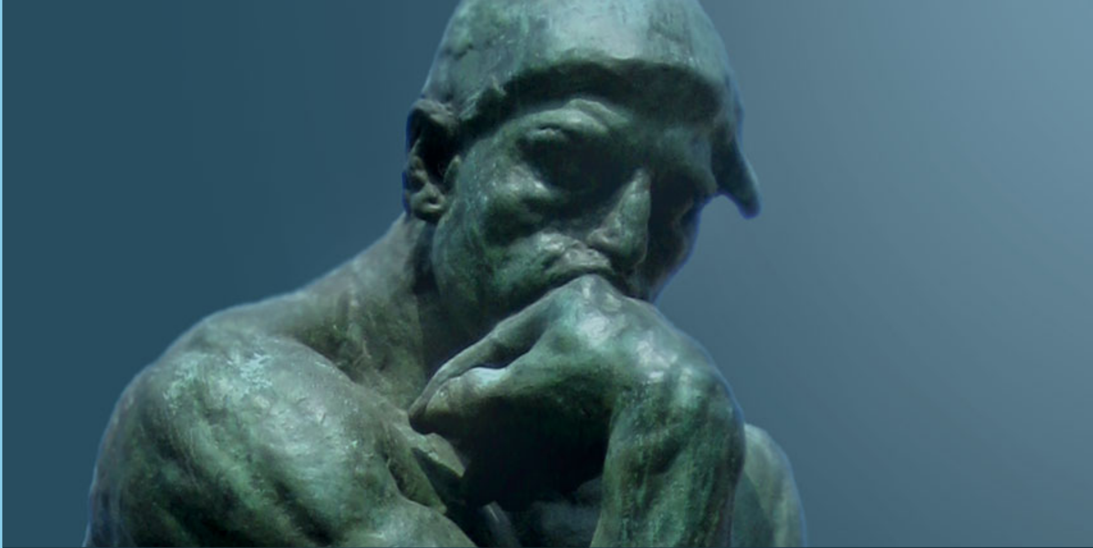


Stephen J. Cowley
Frédéric Vallée-Tourangeau *Editors*

Cognition Beyond the Brain

Computation, Interactivity and Human Artifice

Springer

Cognition Beyond the Brain

Stephen J. Cowley • Frédéric Vallée-Tourangeau
Editors

Cognition Beyond the Brain

Computation, Interactivity and Human Artifice

Springer

Editors
Stephen J. Cowley
School of Psychology
University of Hertfordshire
Hatfield, Hertfordshire, UK
and
Department of Language and Communication
University of Southern Denmark
Slagelse, Denmark

Frédéric Vallée-Tourangeau
Department of Psychology
Kingston University
Kingston upon Thames, Surrey, UK

ISBN 978-1-4471-5124-1 ISBN 978-1-4471-5125-8 (eBook)
DOI 10.1007/978-1-4471-5125-8
Springer London Heidelberg New York Dordrecht

Library of Congress Control Number: 2013941878

Cover image: Auguste Rodin's 'Le Penseur', photo courtesy of Gertjan R. (CC BY-SA 3.0)

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

1 **Human Thinking Beyond the Brain** 1
Frédéric Vallée-Tourangeau and Stephen J. Cowley

2 **Human Agency and the Resources of Reason** 13
Martin Neumann and Stephen J. Cowley

3 **Judgement Aggregation and Distributed Thinking** 31
Kai Spiekermann

4 **Computer-Mediated Trust in Self-interested Expert Recommendations** 53
Jonathan Ben-Naim, Jean-François Bonnefon, Andreas Herzig, Sylvie Leblois, and Emiliano Lorini

5 **Living as Languaging: Distributed Knowledge in Living Beings** . . . 71
Anton Markoš, Jana Švorcová, and Josef Lhotský

6 **The Quick and the Dead: On Temporality and Human Agency** . . . 93
Bent Holshagen Hemmingsen

7 **You Want a Piece of Me? Paying Your Dues and Getting Your Due in a Distributed World** 113
Peter E. Jones

8 **Distributed Cognition at the Crime Scene** 131
Chris Baber

9 **Socially Distributed Cognition in Loosely Coupled Systems** 147
Mark Perry

10 **Thinking with External Representations** 171
David Kirsh

11 **Human Interactivity: Problem-Solving, Solution-Probing and Verbal Patterns in the Wild** 195
Sune Vork Steffensen

12 Interactivity and Embodied Cues in Problem Solving, Learning and Insight: Further Contributions to a "Theory of Hints" 223
Linden J. Ball and Damien Litchfield

13 Naturalising Problem Solving 241
Frédéric Vallée-Tourangeau and Gaëlle Villejoubert

14 Systemic Cognition: Human Artifice in Life and Language 255
Stephen J. Cowley and Frédéric Vallée-Tourangeau

Author Index 275

Subject Index 283

Contributors

Chris Baber Electronic, Electrical and Computer Engineering, The University of Birmingham, Birmingham, UK

Linden J. Ball School of Psychology, University of Central Lancashire, Preston, UK

Jonathan Ben-Naim CNRS and Université de Toulouse, Toulouse, France; IRIT, Université Paul Sabatier, Toulouse, France

Jean-François Bonnefon CNRS and Université de Toulouse, Toulouse, France; CLLE, Maison de la Recherche, Toulouse, France

Stephen J. Cowley School of Psychology, University of Hertfordshire, Hatfield, Hertfordshire, UK; Department of Language and Communication, University of Southern Denmark, Slagelse, Denmark

Bent Holshagen Hemmingsen København N, Denmark

Andreas Herzig CNRS and Université de Toulouse, Toulouse, France; IRIT, Université Paul Sabatier, Toulouse, France

Peter E. Jones Department of Humanities, Sheffield Hallam University, Sheffield, UK

David Kirsh Department of Cognitive Science, University of California, San Diego, La Jolla, CA, USA

Sylvie Leblois CNRS and Université de Toulouse, Toulouse, France; CLLE, Maison de la Recherche, Toulouse, France

Josef Lhotský Faculty of Science, Department of Philosophy and History of Science, Charles University Prague, Praha 2, Czech Republic

Damien Litchfield Department of Psychology, Edge Hill University, Ormskirk, Lancashire, UK

Emiliano Lorini CNRS and Université de Toulouse, Toulouse, France; IRIT, Université Paul Sabatier, Toulouse, France

Anton Markoš Faculty of Science, Department of Philosophy and History of Science, Charles University Prague, Praha 2, Czech Republic

Martin Neumann University of Koblenz, Koblenz, Germany

Mark Perry Department of Information Systems and Computing, Brunel University, Uxbridge, Middlesex, UK

Kai Spiekermann Department of Government, London School of Economics, London, UK

Sune Vork Steffensen Centre for Human Interactivity, Institute of Language and Communication, University of Southern Denmark, Odense M, Denmark

Jana Švorcová Faculty of Science, Department of Philosophy and History of Science, Charles University Prague, Praha 2, Czech Republic

Frédéric Vallée-Tourangeau Department of Psychology, Kingston University, Kingston upon Thames, Surrey, UK

Gaëlle Villejoubert Department of Psychology, Kingston University, Kingston upon Thames, Surrey, UK

Chapter 1
Human Thinking Beyond the Brain

Frédéric Vallée-Tourangeau and Stephen J. Cowley

Abstract It was long assumed that thinking goes on 'in the head': indeed, as recently as twenty years ago, many would have regarded it as absurd to examine thinking with reference to events beyond the brain. The chapters in *Cognition Beyond the Brain* adopt a different perspective: In thinking, people use dispositions from both sides of the skull. Readily observed phenomena—including neural activity—constitute the object of *thinking*, which relates conceptually to the construct *'thinking'*. Like all folk concepts, 'thinking' is a second-order construct used to 'explain' observations or, specifically, how action is—and should be—integrated with perception. As attested in each of the chapters, bodies co-orient to cultural resources while using bidirectional coupling. The focus thus falls on what can be learned about thinking by studying world-side activity. The chapters report empirical, observational and theoretical studies of how people use circumstances (and objects) to act alone, in dyads and in groups. People manage and track attention as they integrate speech and action with gestures, gaze and other bodily activity. In making interactivity part of thinking, a broad range and assortments of parts, procedures and modes of operation are invoked.

Thinking in Action

It was long assumed that thinking goes on 'in the head': indeed, as recently as twenty years ago, many would have regarded it as absurd to examine thinking with reference to events beyond the brain. Not only did behaviourists reject this idea

F. Vallée-Tourangeau (✉)
Department of Psychology, Kingston University, Kingston upon Thames, Surrey, KT1 2EE, UK
e-mail: f.vallee-tourangeau@kingston.ac.uk

S.J. Cowley
School of Psychology, University of Hertfordshire, College Lane, Hatfield, Hertfordshire, AL10 9AB, UK

S.J. Cowley
Department of Language and Communication, University of Southern Denmark, Sdr. Stationsvej 28, 4200 Slagelse, Denmark

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_1, © Springer-Verlag London 2013

but when the cognitive (counter) revolution arrived, most were enthralled by models that described task performance in terms of computation. Using what philosophers termed 'functionalism', this legitimised science based on formal models that were implemented on von Neumann machines. Thus, problem solving, linguistic-analysis and making up 3D visual sketches all came to be pictured as occurring 'within' an algorithmic processing system. By the 1990s, however, the climate had changed. It was increasingly recognised that action, perception and attention affect language and thinking. Pursuing this, *Cognition Beyond the Brain* presents studies of how cognitive skills are deployed in a range of complex tasks and activities.

While neurally enabled, cultural and bodily dispositions contribute to human action, people exploit sense-saturated coordination or *interactivity*, a modus operandi based on coordinating with people/objects while orienting to the cultural environment. From this perspective, the heuristic power of symbolic, connectionist, robotic or dynamical models can be separated from normative assumptions. This is done by setting out, not to model hidden states, but to understand how bodies (plural) coordinate. Human interactivity enriches action and collective life by connecting norm-based procedures with the statistical power of information compression. Though people use experience and feeling, they do so in a world where contexts tend to be highly predictable. As a result, a wide range of cultural institutions (e.g., the family, farming, science) exert effects on human brain-body systems. Further, because these evolved to draw on movement and affect, there is no need to reduce thinking to Shannon information that correlates with semantic content. Rather, because thinking uses evaluative functions that have developed over long time scales, formalizations apply to second-order phenomena. Indeed, it is because these are *abstracta* that no one claims that the Internet thinks. Rather, like brains, it serves a human community as a cognitive resource.

In thinking, people use dispositions from both sides of the skull. Readily observed phenomena—including neural activity—constitute the object of *thinking*, which relates conceptually to the construct *'thinking'*. Like all folk concepts, 'thinking' is a second-order construct used to 'explain' observations or, specifically, how action is—and should be—integrated with perception. Our strategy is to seek its basis in adaptive primate behaviour. However, it is also emphasised that, especially when young, human individuals show remarkable flexibility. This appears, above all, in the use they make of language, artefacts and culture. As attested in each of the chapters, bodies co-orient to cultural resources while using bidirectional coupling. The focus thus falls on what can be learned about thinking by studying world-side activity. Though this almost certainly reconfigures plastic brains (e.g., Maguire et al. 2000; Anderson 2010), that is not the topic of the volume. Rather, the chapters report empirical, observational and theoretical studies of how people use circumstances (and objects) to act alone, in dyads and in groups. People manage and track attention as they integrate speech and action with gestures, gaze and other bodily activity. In making interactivity part of thinking, a broad range and assortments of parts, procedures and modes of operation are invoked.

Cognitive Science: The Last 20 Years

Work on robotics and the brain shows that models of how sensory input can be processed by von Neumann machines fail to capture the dynamical complexity of self-organizing biological systems. Yet, aspects of thinking *can* be modelled by algorithms whose strings 'correspond' to semantic content. For example, vision can be described by generating a 3D sketch or, alternatively, designing machines that use pattern recognition to act as if they 'see'. Using Shannon-information produces results that carry functional information for human observers (but not, of course, artificial systems). To suppose that there is a parallel between these models and what brains do is to adopt what Cowley and Spurrett (2003) call an "epistemological conception of mind (p. 290)". Ultimately, Humean and Cartesian traditions led to epistemic focus on disembodied, disembedded representations that can be modelled on alphabetic or imagistic patterns (i.e., associated with invariant 'content'). These can function as 'input' that is decoded and mapped onto physical symbols that are processed by algorithms that generate various kinds of output. As formal models, such approaches have enormous power.

Problems arise from making thinking representational. First, when modelled as information-processing, systems depend on design. Because they are separated from the designer's world, they face the notorious symbol-grounding problem (see Harnad 1990; Belpaeme et al. 2009). This is the converse of what we find in nature. There, complex ecosystems sustain living brains, bodies and people. Indeed, representational models even treat sensory systems as akin to computer peripherals. Further, by the 1990s, it had became abundantly clear that, instead of relying on algorithms, robots could use the world as its own representation. They relied on materials—(nonliving) bodies and environments—and modes of engagement that used non-linear mathematics. Much of nature's complexity depends on flexible brains that adapt as they grow together with moving, perceiving bodies. Unlike programs that use invariant input-patterns, selective processes function to reshape sensorimotor experience.

The concept of 'input' served to rethink behaviourist theory. While Turing treated computation as extending human powers, psychologists tended to conceptualise computation around actual von Neumann machines. By linking these to the epistemic conception of mind, they were able to ignore the objection that biological systems lack equivalents to keyboards. This is because philosophy often assumes that, while animals are automata, human flexibility arises in dealing with propositional attitudes (or similar). In fact, even brainless systems act flexibly: bacteria can choose between sugar solutions (Monod 1942; Shapiro 2011). This is possible because dynamics generate functional information *for* the bacteria. They depend on mechanisms or *effectivities* that cannot be reduced to information processing. These systems arose as ecosystems co-evolved with organisms that realise values or, simply, act adaptively. Of course, once brains evolved, new possibilities came into being. Organisms began to link perception with action and, with learning, to use statistical indicators of environmental valences. In starting with biology, mind ceases to resemble the 'filling' of the input-output sandwich (Hurley 2001). As Turing thought, computation extends, but does not explain, human powers (Wells 2006).

An epistemic conception of mind is thus, as Lyon (2006) suggests, *anthropogenic*. The alternative is to take biology seriously.

Although increasingly based on the study of living systems, the shift to biology began with Varela, Thompson and Rosch's *Embodied Mind* (1991). Second-generation cognitive science focused attention on lived experience. From a first-person perspective, life is embodied and, in some sense, embedded. Further, if one seeks to avoid dealing with reports of experience, one can turn to bodily 'modalities'. Given the ambiguities of appeal to 'embodiment', many focus on how sense-specific histories contribute to later action and perception. On such an approach, cognition is seen as grounded (Barsalou 2010): humans become multimodal situated agents. Rather than debate the value of connectionist, robotic or dynamical models, weight falls on how tasks are accomplished. Neuroscience is thus increasingly complemented by work on how organisms (including people) act in the world (Thompson 2007; Robbins and Aydede 2009; Chemero 2009; Stewart et al. 2010). Moving beyond the negative claim that cognition is not brain-bound, new debates have flourished. On the one side, many propose embedded and/or extended functionalism; on the other, another grouping build on *The Embodied Mind* to propose that cognitive science adopt the enactivist paradigm.

Philosophy is plagued by the so-called mind-body problem. In first-generation cognitive science, this was avoided by positing that intelligent behaviour depends on *mental states*. On a functionalist view, these states (or systems) play a causal role. They can, for example, link a physical injury (input) to the belief that one is in pain (or, rather, a representation of that belief). While traditionally based on input, the model can include the body and, in extended mind, external vehicles. Today functionalists debate whether brains alone carry 'the mark of the cognitive' (Adams and Aizawa 2010). Offering an alternative, Clark and Chalmers (1998) influentially argued that artefacts extend the mind. Using the parity principle they suggest that if external vehicles have the same functional role as mental states, they too serve as part of the cognitive process. However, both sides place the organism at the core of cognitive systems. Indeed, Clark (2008) views this as uncontroversial. Rather than scrutinise how organisms use the environment to anticipate what may happen, the brain is seen as a predictive engine. Further, mechanism is closely identified with function—no attempt is made to decompose these into parts, operations and modes of organization. Functionalists thus retain the classic view that language is a verbal system that serves, among other things, to specify propositions. Like a phonetic alphabet, language consists in symbols whose structure appears in analysis. Written language—and translation—show that language can be seen as a system unto itself—a means of mapping forms onto meanings (and vice versa). On the extended mind view, brains are able to learn the properties of material symbols; for active internalists, as Dennett (1991) phrases it, they 'install' a virtual language machine. To the extent that thinking draws on words, therefore, depends on plugging beliefs into how language is 'used' in accordance with grammar and logic.

Others emphasise phenomenological experience. While Searle's (1980) Chinese Room thought experiment challenged functionalism to explain how consciousness arose in biology, more recent work looks elsewhere. Instead of treating mental states (or representations) as biological, the enactivist seeks a common basis for first and third person views. Living systems are taken to possess the functionality that allows cells to engage with the world (though structural coupling), maintain self-organization (by means of internal reorganization) and, where necessary, deal with change (though adaptivity). The enactivist paradigm for cognitive science (see Stewart et al. 2010), has used artificial life in building cognitive models. Their successes mean that discussion of representation and vehicles is giving way to interest in how sensorimotor activity links perception-action with neural function. Recently, this has been linked to participatory sense-making (De Jaegher and Di Paolo 2007), a form of social behaviour whose complexity resembles that of bacterial communication/cognition (emergent patterns of interaction between agents affect decisions—and trigger unforeseen consequences). In experiments on what is known as perceptual crossing (Auvray et al. 2009) participants engage with three identical stimuli of which one is controlled by a person. Crucially, using interactivity, they identify this more often than would be the case by chance. Cognition connects material entities with how agents orient to each other in what is called a consensual domain (Maturana 1978). However, it is also possible to allow activity to reiterate linguistic patterns by drawing on motor movements or phonetic gesture. On this view, as people learn to speak, they discover the effects that arise from what Bottineau (2012) calls linguistic techniques.

Much work in cognitive science fits neither of these categories. Using cognitive ethnography, Hutchins (1995) proposed functional models of how individuals and groups use artefacts during complex tasks (e.g., in landing a plane). In this, he treated representations as entities that are propagated in public. While often seen as a variant of extended mind, the approach is more radical. This is because Hutchins's public representations link experience with physical patterns. Far from reducing to material symbols, they are embedded in cultural process. The radicalism comes to the fore when one recognises that non-linguistic experience is bound up with wordings. Thus, while part of action, language *is* also part of history. This insight shapes a view where, in Love's (2004) terms, second-order cultural constructs or verbal patterns ('words') are perceived as part of first-order activity (or action-perception). During talk, people draw on interactivity to create and construe wordings. First-order language is thus measurable whole-bodied activity that, oddly, evokes second-order patterns (including 'meanings'). Full-bodied metabolic activity therefore enacts sociocultural patterns. The resulting distributed view of language thus blends with ecological psychology and Chemero's (2009) 'radical embodied cognitive science.' For Cowley and Vallée-Tourangeau (2010), this 'more subtle' challenge to the epistemic view of mind builds, in Hollan et al.'s (2000) terms, on how the products of earlier (cultural) events transform later events. In linking neural function and with the slow dynamics of linguistic and cultural change, interactivity makes human cognition central to how people live temporal experience.

Systemic Cognition: Making a Start

Cognition Beyond the Brain had its beginnings at a symposium on 'Distributed thinking' in Hertfordshire in 2009. By connecting thinking with action, participants addressed how cultural organization influences people's dealings with both objects and each other. In presenting some of the resulting papers in a special issue of *AI and Society*, Cowley and Vallée-Tourangeau (2010), emphasised that people often manage cognitive tasks by drawing on co-acting assemblages. We hoped that identification of such assemblages could be used to place bounds on *thinking in action*. The chapters by Baber, Perry, Ben-Naim et al., Spiekermann and Jones all began at the Distributed Thinking Symposium and can be read as describing how cognition spreads beyond the brain. However, a striking problem arose. Above all, the work on computer-mediated trust (see, Ben-Naim et al. 2013, Chap. 4) shows that an assemblage based view of distributed cognition reaches far beyond the domain of 'thinking'. Baber and Jones thus concur that the approach loses sight of what is truly cognitive: In Jones's (2013, Chap. 7) terms, the only *active* part of a cognitive system is a living human being. Accordingly, in a later symposium at Kingston University ('Rethinking problem solving'), new importance was given to questions of human agency. Building on this meeting, Ball and Litchfield (2013, Chap. 12) present evidence supportive of the view, first argued by Kirsh, that interactivity is central to cognition. They show that, in reading X-rays, novices draw on how the image guides real time saccading. They connect what can be seen with the coupling of organism-environment relations. In development, Neumann and Cowley (2013, Chap. 2) argue, similar processes result in allowing people (and their brains) to make use of cultural resources. As individual agents become increasingly rational, they learn to participate in a cognitive-cultural network. Brains and genes function in an extended ecology where interactivity contributes to learning to talk and, indeed, prompts discovery of how to seek, identify and solve problems. In placing emphasis on how humans contribute to co-acting assemblages, emphasis fell on interbodily activity (i.e., ways of engaging with objects and people). Having established its importance beyond the skin, we elicited a paper that made comparison with molecular processes (Markoš et al. 2013, Chap. 5) and one on how a brain-damaged person acts to construct the *now* (Hemmingsen 2013, Chap. 6). Finally, new papers explore the role of interactivity in language (Steffensen 2013, Chap. 11) and in problem solving (Vallée-Tourangeau and Villejoubert 2013, Chap. 13).

Kirsh's (2013, Chap. 10) 'Thinking with External Representations' shows how interactivity opens up the previously unthinkable. This depends on the remarkably large number of ways in which even simple physical entities can be used as affordances. While they can be seen as a stable form of output or fixed patterns, they also serve to set off incremental processes that create (shareable) thoughts. They promote human projecting and materialising that drive both efficiency/effectiveness and other cognitive processes. Vallée-Tourangeau and Villejoubert cash out Kirsh's (2013) dictum "there is magic in actually doing something in the world" by illustrating in a range of laboratory-based problem solving tasks that interactivity with a modifiable problem presentation produces unanticipated paths to solution. This augmented creativity defuses mental set and facilitates representation restructuring in overcoming impasse in insight problem solving.

Exploring related ideas in a professional setting, Steffensen's (2013) cognitive event analysis offers a rich description of problem solving 'in the wild'. He shows that what people do and say depends heavily on very rapid events. What he terms 'insight' arises as full-bodied activity, and is concerted in the pico-scale of tens of milliseconds. Like problem-solving, language arises as action and perception are managed. Interactivity thus becomes an 'ontological substrate' for language and cognition (Steffensen 2013, Chap. 11). Much depends on how people coordinate and, as a result, manage events (and each other) in different time scales.

Baber's (2013, Chap. 8) paper on 'Distributed Cognition at the Crime Scene' emphasises change over time. While investigation begins by using various resources to create narratives—as gaze, for example, shapes interactivity, there are gradual changes in the kinds of thinking that are required. In developing wearable technology, there is a need to give 'structure' to what occurs in court. Captured affordances, evidence, must sustain objective judgements based on individual 'mental activity': implicitly, early on decisions emerge and, later, decisions must be made. Cowley and Vallée-Tourangeau (2013, Chap. 14) argue that, like language, thinking has a strange duality in that it too is grounded in activity while managed in line with cultural and organizational constraints. Paradoxically, on the systemic view, certain 'forms of cognition' (Loughlin 2012) depend on how individuals draw on available linguistic and institutional resources.

Ball and Litchfield (2013) show much the same for expertise. They demonstrate how interactivity is deeply implicated in gaze. Far from relying on physical invariances, people exploit cues during goal directed behaviour. Like heuristics, these extend adaptive processes. Further while cues can function consciously—-as when a pulsating image invites conclusions—they also function in more subtle ways. Hinting at a transfer of 'intentionality', they show that the performance of novices who interpret a lung x-ray improve when they are covertly cued by expert gaze. While calling this grounded cognition, simulation of sensorimotor action and input-output processes, Ball and Litchfield (2013) suggest that people draw on external resources by using sensorimotor control to decide how to make the most of what can be seen.

Computation, Interactivity and Human Artifice

On a systemic view, the focus shifts from modelling hidden mechanisms to the investigation of how results (Järvilehto 2009), or the products of the organism-environment system, are obtained. Results "embod(y) both preceding organization of the system and its potential for future behaviour and future results" (Järvilehto 2009, p. 116). The volume presents much empirical work showing that action and perception direct attention to structures whose origins lie in both biological and cultural time-scales. While brains contribute to cognitive tasks, over time, people learn to use cultural products to affect later action and perception. As Järvilehto stresses, *joining organs* prompt people to anticipate what is likely to occur not only immediately but also in longer scales. Typically, interbodily activity fine-tunes how people engage with each other, artefacts and how language links embodiment with verbal patterns. Although the results can often be explained (or explained away) by the concept of "thinking", they appear in observable activity. Indeed, we may come to *believe* in mind and languages through linking a history of acting and perceiving to folk explanations of our doings. In learning to conform/ strategise, or act in rational ways, we discover a partly shared world. Where one understands other peoples' perspectives, much can be gained by orienting to likely judgements. While this bears on debate between functionalists and enactivists (Cowley and Vallée-Tourangeau 2013), in this context we focus on simpler matters.

Sixty years of computational and robotic modelling yielded modest insight as to how individuals think and perceive. Appeal to Shannon-information and statistics fails to clarify how, in practice, we implement thinking and language. This is because all such models overlook functional information. Living systems depend on, not input (or pure statistics), but bidirectional coupling. For an astute observer, even bacteria choose what to do. On the other hand, computation comes into its own in seeking to understand more abstract patterns (e.g., numbers, syntax). As Turing (1936) thought, it may function by extending cognition (not 'explaining' it). Populations, not individuals, use functional (Shannon) information. Not only does this appear in computer generated trust and judgement aggregation but also in second-order aspects of language. Though bound up with action/perception, human populations use verbal patterns with striking predictability. In the extended ecology, eusocial primates draw on a selection history that compresses functional information and, by so doing, makes it increasingly Shannon-like. In the domain of language, at least, these resources allow them to bring their doings under a degree of collective control.

Human encounters with the world are embodied and embedded or, more precisely, based in sense-saturated coordination. Interactivity matters to human action and perception. Even if bidirectional coupling is (almost) as old as life itself, humans use felt activity to manage attention and perception. They link bodily encounters with experience of what is likely to happen between people: much thinking has an interbodily basis. As people perceive, they act and, as they act, they perceive: in Gibson's (1979) sense, learning is the education of attention. Gradually, people discover the duality of language—they shape a flow of first-order activity while drawing on a background of second-order patterns. Using reiterated activity, the phenomena of decision-making are automatized. To do this, however, people need new forms of cognitive control. Individuals develop a sense of how circumstances can be used as they develop a singular identity. People depend on traditions, practices and wordings—modes of life where circumstances evoke cultural products that have the power to alter later activity.

The capacity to link circumstances with a flow of acting, feeling and thinking is often called *creative*. To our knowledge, the only established alternative is Peirce's (1891) objective idealism. However, *Cognition Beyond the Brain* hints at another view. Novel thoughts and actions are traced to changes in the flow of action. Often, people use the products of past events to set up what Steffensen (2013) calls problem seeking and solution probing. While not goal-directed, this behaviour enables people to coordinate in ways that realise values. They enact reiterated activity but not, of course, perfect repetition: as Heraclitus saw, there is only flux. Accordingly, they satisfice (or improvise). Perhaps this is because brains work on the edge of chaos or, in Wallot and Van Orden's (2011) terms, in states of continuous frustration. Sometimes felt movement produces valued results; sometimes these arise from inhibition. This may seem miraculous. Of course, the basis lies in experience of using interactivity while orienting to second-order constructs that are evoked by circumstances. If we are correct, human thinking is inextricable from the history of human artifice.

References

Adams, F., & Aizawa, K. (2010). Defending the bounds of cognition. In R. Menary (Ed.), *The extended mind* (pp. 67–80). Cambridge: MIT Press.

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33*, 245–313.

Auvray, M., Lenay, C., & Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, *27*, 32–47.

Baber, C. (2013). Distributed cognition at the crime scene. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Ball, L. J., & Litchfield, D. (2013). Interactivity and embodied cues in problem solving and insight: further contributions to a "theory of hints". In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Barsalou, L. W. (2010). Grounded cognition: past, present and future. *Topics in Cognitive Science*, *2*, 716–724.

Ben-Naim, J., Bonnefon, J. F., Herzig, A., Leblois, S., & Lorini, E. (2013). Computer-mediated trust in self-interested expert recommendation. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Belpaeme, T., Cowley, S. J., & MacDorman, K. F. (2009). *Symbol grounding*. Amsterdam: Benjamins.

Bottineau, D. (2012). Remembering voice past: languaging as an embodied interactive cognitive technique. In *Proceedings of the humanities international forum*. Moscow: Russian State University for the Humanities.

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge: MIT Press.

Clark, A. (2008). *Supersizing the mind: embodiment, action and cognitive extension*. Oxford: Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7–19.

Cowley, S. J., & Spurrett, D. (2003). Putting apes (body and language) together again. *Language Sciences*, *25*, 289–318.

Cowley, S. J., & Vallée-Tourangeau, F. (2010). Thinking in action. *AI & Society*, *25*, 469–475.

Cowley, S. J., & Vallée-Tourangeau, F. (2013). Systemic cognition: human artifice in life and language. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: an enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*, 485–507.

Dennett, D. (1991). *Consciousness explained*. Boston: Brown.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Harnad, S. (1990). The symbol grounding problem. *Physica D. Nonlinear Phenomena*, *42*, 335–346.

Hemmingsen, B. (2013). The quick and the dead: on temporality and human agency. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, *7*, 174–196.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, *19*, 265–288.

Hurley, S. L. (2001). Perception and action: alternate views. *Synthese*, *129*, 3–40.

Järvilehto, T. (2009). The theory of the organism-environment system as a basis of experimental work in psychology. *Ecological Psychology*, *21*, 112–120.

Jones, P. (2013). You want a piece of me? Paying your dues and getting your due in a distributed world. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Kirsh, D. (2013). Thinking with external representations. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Loughlin, V. (2012). Mark Rowlands, The new science of the mind: from extended mind to embodied phenomenology. *Phenomenology and the Cognitive Sciences*. doi:10.1007/s11097-012-9274-y.

Love, N. (2004). Cognition and the language myth. *Language Sciences*, *26*, 525–544.

Lyon, P. (2006). The biogenic approach to cognition. *Cognitive Processing*, *7*, 11–29.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural changes in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 4398–4403.

Markoš, A., Švorcová, J., & Lhotský, J. (2013). Living as languaging: distributed knowledge in living beings. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Maturana, H. (1978). Cognition. In P. Hejl, K. Wolfram, K. Köck, & G. Roth (Eds.), *Wahrnehmung und Kommunikation* (pp. 29–49). Frankfurt: Peter Lang.

Monod, J. (1942). *Recherches sur la croissance des cultures bactériennes*. Paris: Hermann.

Neumann, M., & Cowley, S. J. (2013). Human agency and the resources of reason. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Peirce, C. S. (1891). The architecture of theories. *The Monist*, *1*, 161–176.

Robbins, P. & Aydede, M. (Eds.) (2009). *The Cambridge handbook of situated cognition*. Cambridge: Cambridge University Press.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*, 417–457.

Shapiro, J. A. (2011). *Evolution: a view from the 21st century*. Upper Saddle River: FT Press.

Steffensen, S. V. (2013). Human interactivity: problem-solving, solution-probing and verbal patterns in the wild. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Stewart, J., Gapenne, O., & Di Paolo, E. A. (Eds.) (2010). *Enaction: toward a new paradigm for cognitive science*. Cambridge: MIT Press.

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

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society (2)*, *42*, 230–265.

Vallée-Tourangeau, F., & Villejoubert, G. (2013). Naturalising problem solving. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

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

Wallot, S., & Van Orden, G. (2011). Grounding language performance in the anticipatory dynamics of the body. *Ecological Psychology*, *23*, 157–184.

Wells, A. (2006). *Rethinking cognitive computation: Turing and the science of the mind*. Basingstoke: Palgrave Macmillan.

Chapter 2
Human Agency and the Resources of Reason

Martin Neumann and Stephen J. Cowley

Abstract The evidence shows that human primates become (relatively) rational actors. Using a distributed perspective, we identify aspects of human agency that raise important questions for sociocognitive science. In humans, we argue, agency does not centre on individual agents. Cognitive, social and linguistic outcomes depend on skills in moving in and out of aggregates that bind people, artifacts, language and institutions. While recognising the value of symbol processing models, these presuppose the embodied events of *human symbol grounding*. At a micro level, humans coordinate with others and the world to *self-construct* by cognising, talking and orienting to social affordances. We trace the necessary skills to sense-saturated coordination or *interactivity*. As a result of perceiving and acting on the environment, human individuals use the artificial to extend their natural powers. By using verbal patterns, artifacts and institutions, we become imperfect rational actors whose lives span the micro and the macro worlds.

Towards Sociocognitive Science

Humans make extensive use of cultural resources that include languages. Alone of the primates, their infants develop in a normative world that allows species to develop a distinctive type of agency. This is manifest individually and, just as strikingly, in how human groups implement long-term plans. For historical reasons, attempts to explain human agency have typically sought its origins in, not a history of

M. Neumann
University of Koblenz, Koblenz, Germany
e-mail: maneumann@uni-koblenz.de

S.J. Cowley (✉)
School of Psychology, University of Hertfordshire, College Lane, Hatfield, Hertfordshire, AL10 9AB, UK
e-mail: s.j.cowley@herts.ac.uk

S.J. Cowley
Department of Language and Communication, University of Southern Denmark, Sdr. Stationsvej 28, 4200 Slagelse, Denmark

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_2, © Springer-Verlag London 2013

acting in the world, but genes and/or brains. While biological factors are crucial, we argue that their function is to allow living beings to exploit the environment to develop cognitive, social and linguistic skills. The process arises as infants learn to talk or, in exceptional cases, make use of signing. We therefore argue that it is through participating in language-activity that human primates *become* rational actors. Ours is thus a reworking of what makes us human in terms of Durkheim's (1895 [1998]) famous claim that only the social can ground the social.

Embodied coordination enables infants to self-construct as members of a culturally extended ecology. Once we recognise that social predispositions for embodied coordination are functionally reorganized by encultured body-world activity, Durkheim's view ceases to appear circular. Social behaviour arises as we are moved by others to coordinate in increasingly complex ways. By hypothesis, bodies self-organise by learning to cognise, speak and act strategically: human agency develops within aggregates that bind together artifacts, institutions, and ways of experiencing the world. It is a human capacity for moving in and out of such aggregates—for exploiting embodied coordination—that enables social action to drive the self-construction of rational human actors. In tracing human agency to cognition beyond the brain, a history of coordination is seen as the basis of knowledge. Infants use spontaneous activity in learning to orient to a population's linguistic and other social practices. On the distributed perspective, this shapes human agency—infants discover ways of achieving effects that are, inseparably, cognitive, linguistic and sociological.

The Distributed Perspective

Human agency has previously been traced to how acting and thinking are honed by the demands of sociocultural environments. This is done in, for example, activity theory, cultural psychology and the pragmatism of Mead. What is novel to the distributed perspective is the view that agency results from acting and perceiving in socially distributed systems. For readers of *Cognition Beyond the Brain*, the idea will be familiar. The agency of a pilot who lands a plane is non-local in that, as Hutchins (1995) shows, it is distributed across bodies (and brains) that coordinate an aggregate of resources. The pilot uses readings from the instrument panel, messages from ground control, and real time interaction with a co-pilot. Far from centering on a neural system, agency arises in *acting* with material, cultural, and social structures. Ethnography makes that clear. To understand how such aggregates function, however, systemic output must be separated from construals and associated actions. In Tetris, human-computer aggregates rely on both goal-directed actions and epistemic actions that use the world to simplify cognitive tasks (Kirsh and Maglio 1994). Epistemic actions depend on sense-saturated coordination or player-computer *interactivity*. They change on-screen resources in ways that suit the player's expert knowledge. In conversations, tight coupling allows people to concert by using, not just what is said, but also how speech and movements are integrated in rapid or pico time-scales.[1] In tight coupling that uses cultural resources, in Cronin's (2004) terms, we are cognitively cantilevered into the *Umwelt*. Coordination that is faster than conscious perception drives spontaneity by linking expertise with the results of joint action. The thesis of our paper is that this sense-saturated interactivity shapes human agency by giving us skills in dealing with artifacts, people, and languages.

In ethnography, language is identified with the words that are actually spoken (and which can be transcribed). However, its functionality depends on the fact that language too is *distributed* (see, Cowley 2007c, 2011a; Thibault 2011). In its paradigmatic face-to-face settings, language contributes to action by social aggregates: it is activity in which human beings re-enact the cultural practices and patterns of the ecology. Since it is both situated and verbal, people draw on each other's thinking. This is possible because, unlike animal signalling, languaging has a non-local aspect. Embodied activity links circumstances to past situations by virtue of how we perceive verbal patterns. During languaging, cultural constraints prompt real-time construal. In conversations, verbal patterns arise as concerted acts of articulation (or signing) are accompanied by facial and other expression. Since the results are both embodied and non-local, infants can link metabolic dynamics with between-people coupling. Without hearing what is said, interactivity—and the feeling of thinking—sensitise them to normative resources. While initially reliant on circumstances, they gradually come to hear language as consisting in thing-like parts or to *take a language stance* (Cowley 2011b). Once utterances sound like utterances of something, perceived results or *wordings* can be used in, for example, asking about things. Language, agency and rationality are, irreducibly, individual *and* collective. Cognition links the world in the head with the physical, linguistic and cultural processes of an extended ecology (Steffensen 2011)—a place where individual actions carry cultural value. Human agency self-organises during individual development. Infants use circumstances to become intentional and, later, make use of the resources of reason. Since language is ecological, dialogical and (partly) non-local, rationality co-emerges with individual agency. Although based in interactivity, ways of feeling out the world are supplemented by intelligent—partly conformist—use of external resources.

Agency and Human Agency

The term *agency* can be applied to people, animals, social institutions and inorganic processes. In the first place, it is therefore useful to distinguish physico-chemical agents and living systems. Only in biology do systems (and lineages of systems) set parameters (see, Pattee 1969, 1997) that allow them to measure and control aspects of their environments. Living systems are adaptive and yet also able to maintain autonomy (see, Di Paolo 2005). Adaptive self-organisation allows even single celled

[1] This is the time-scale within which gestures are made and syllables articulated. It can broadly be associated with a window of around 200 ms. It is especially important in prosody (see, Cowley 2009) but, at this scale, interactivity is full-bodied (for detailed discussion, see Thibault 2011).

bacteria to explore their worlds by linking genes and metabolism with viable use of available resources. Flexibility increases in living-systems that use brains, development and learning. Yet, these processes too depend on self-organisation or how organic systems exploit the world beyond the body (including other organisms; Thompson 2007). Organisms are aggregated systems whose parameters link a lineage's history with, in embrained species, experiential history of encountering the world.

In evolutionary time, organisms show flexibility as they adapt to the world and, more strikingly, adapt the world to them. For Jarvilehto (1998, 2009), the world of the living is to be conceptualised in terms of interlocking Organism-Environment Systems (O-E systems). The necessity of the view appears with perceptual learning: as Raguso and Willis (2002) show, foraging hawk-moths learn about local sets of flora. In optimising their behaviour, their partly plastic brains link genetically based self-organisation with learning about an environment. Further, as conditions change, they alter the parameters. The example is apposite in that, while such intelligence is organism-centred, this does not apply to all insects of comparable neural dimensions. In bees and other eusocial insects, cognition serves the colony rather than individuals. Below we argue that languages, technology and money make humans partly eusocial. Since we live in a culturally extended ecology, we are *hypersocial* beings (Ross 2007) whose primate intelligence is extended as we orient to eusocial resources that function as cultural (second-order) constraints. Individual-environment relations thus transform individual experience, learning and ontogeny.

Human uniqueness depends on neither our hypersocial nature nor our propensity to exploit the world beyond the brain. What is most striking about human agency is how it combines artificial rigidity with our natural flexibility. As organism-environment systems, we detect *rationality*; as populations, we tend to act in line with utility calculations. Uniquely, humans are partly biological and partly rational. As individuals, we grasp rules (imperfectly), ascribe minds to agents, plan, take part in social institutions and make use of wordings, tools and machines. However, we draw on the resources and skills of populations. How is this to be explained? While bound up with learning to talk (not to mention literacy and numeracy), human agency also uses artifacts and institutions. These perform a dual function both as boundary conditions and as flexible constraints: they serve to measure and also to control. Given the relative predictablity of wordings, we extend our natural intelligence. Accordingly, we now turn to how coordination alters a social actor's sensorimotor and cultural strategies. Coming to act in line with utility calculation depends on learning to concert movements with those of others, exploit available social strategies and, using these, gaining skills in using the artifacts and cultural resources of a community.

Language and Languaging

Since the 18th century human nature has been associated with a mental locus of *ideas*. On such a view, language becomes a transparent conduit between minds

(Locke 1690 [1975]) used to construe verbally-based thoughts (see, Reddy 1979). In the 19th century, this set of metaphors froze as a *code view* that gave us, first, Morse and telegraphy and, later, computers and the Internet. Given the influence of technology, these metaphors were taken up by Saussure and subsequently dominated twentieth century linguistics. However, following Harris (1981), Linell (2005), Love (2004), Kravchenko (2007) and Bickhard (2007) a growing number reject *encodingism*. Far from being a conduit of ideas that are encoded/decoded by minds or brains, language is an ecological whole that functions in many time-scales. It is metabolic or dynamical and, at once, symbolic (Rączaszek-Leonardi 2009). Though part of concerted activity by at least one person, its products are, at once, developmental, historical and evolutionary. Wordings are enacted and, yet, use traditions that are constitutive of the social world. Computers—not living beings—rely on symbolic processes function to encode/decode physical states. Thinking is action: on a machine analogy, total language trains up networks of hypersocial cultural agents that, in their lifetimes, attempt to 'run' languages.

On the conduit view, so-called language 'use' is said to result from the workings of language-systems (e.g., isiZulu, English). Its basis is ascribed to individual knowledge that is represented by a mind or brain. As in Western grammatical tradition, language is described—not around observables (i.e., articulatory activity and pulses of sound)—but by relations between phenomenological forms. Language is thus identified with words, grammars, discourse or constructs that, in some mysterious way, 'reflect' inner thought. Like an artificial system, a brain maps forms onto meaning and, conversely, meanings onto form. Among the problems with any such view is the mereological error of supposing that 'forms' serve brains as input or output. Rather, *people* make and track phonetic gestures that shape how they perceive speech. However, there are no determinate forms *in* the speech wave and we rely on how things are said. As an avalanche of evidence shows, brains exploit rich phonetic information (for review, see Port 2010). Further, we find it hard to track the precise sense of *the words that are actually spoken*. Since connotations affect construal, why does the code view of language persist? Leaving aside the sociology of science, there are two reasons. First, we rely on *the language myth* (Harris 1981): in everyday life and many institutions language is conceptualised in terms of forms that 'contain' messages. Second, inscriptions use writing systems that invite us to think that form is its essence: since language can be reformatted, we view writing as 'like' language—its essence lies in potential reformatting. Many fall prey to what Linell (2005) calls *written language bias*. It is forgotten that, in themselves, inscriptions lack meaning. Like meanings, forms are *abstracta*.

Later, we adduce further reasons for rejecting code views. However, one of the most compelling is that these reify the phenomenological. True to orthodox science, we prefer to begin with measurable phenomena. By starting with speech coordination we commit ourselves to addressing the goal of how this comes to be *perceived* in terms of forms and meanings. Rather than posit a mental or neural locus, this depends on how language spreads across populations. Perception of form and meaning is a multi-agent phenomenon and, thus, language is distributed (see, Cowley 2011a). It is therefore important to distinguish *languaging* from its products (vocal and other gestures) and the related phenomenology (wordings). Languaging is full-bodied, dialogical, situated and amenable to measurement. Provisionally, it can be defined as "*face-to-face routine activity in which wordings play a part*". Rather than treating forms or meanings as primary, it is recognised that we perceive bodily events as wordings.[2] Emphasis on coordination allows due weight to be given to the fact that languaging predates literacy by tens-of-thousands or years. By hypothesis, all linguistic skills derive from face-to-face activity or languaging. However, it is only over time that children come to make use of these phenomenologically salient and repeatable aspects of second-order cultural constructs (Love 2004). Given the many ways in which they contribute to languaging, they have meaning potential that gives language with a verbal aspect. As Thibault (2011) points out, linguists typically confuse language with second-order constructs. Importantly, in making wordings second-order, we contrast their ontology with that of languaging. Pursuing the contrast, we can use an analogy. Gibson (1979) compared perceiving the world with perceiving pictures. On a card showing Rorschach dots we may see an arrangement of markings and, for example, a dancing bear. Using *discrepant awareness*, we pick up both invariants *of* the picture (e.g., the dots) and invariants *in* the picture (the 'bear'). In languaging too, we pick up invariants of the activity (e.g., how people speak, gesture and use their faces) as well as invariants *in* the activity (e.g., wordings and meanings). Like learning to see pictures, learning to talk draws on discrepant awareness. Just as we see 'things' in pictures, we hear 'things' in utterances. In Cowley's (2011b) terms, we *take a language stance*. On the analogy, this is like learning to see things in pictures while, at the same time, using the body to make one's own (verbal) images. And that, of course, presupposes human agency. Next, we turn to how infants exploit languaging—activity in which wordings play a part—to self-organise their bodies and become human agents who perceive—and make up—wordings.

Human Symbol Grounding

To become fully human, children have to discover how to behave and, among other things, learn how wordings contribute to collective practices. As they become able to play various roles, they benefit from acting and speaking in particular ways. Initially, learning to talk depends on managing concerted action—interactivity—just as much as on wordings. Unlike symbol processors, we use circumstances in coordinating in ways likely to achieve strategic ends. At times we act as others expect and, thus, make what count as valid judgements. Practical skills and shared knowledge shape social action. This, of course, connects ontogenesis, training and education. Next, therefore, we sketch how infants use human symbol grounding to

[2] Adults sing, converse, read books, discover new media, and are fooled by advertisements: wordings appear in dreams and silent thoughts. While not languaging, this is also what Love (2004) calls *first-order language*. In all of these activities formal patterns can be said to constrain expressive dynamics.

sensitise to wordings. Whereas babies rely on interactivity, by 3–4 years of age, second-order constructs (and wordings) exert tight constraints on how children act, think and feel. In tracing how the symbols of a community become part of a person, we depend on what Cowley (2007a) terms *human symbol grounding*.

Though the symbols to be grounded consist in more than wordings, these are foundational.[3] Because they are jointly enacted, they link statistical learning, norms and first-person phenomenology. This triple process begins as a baby's brain sensitise to cultural routines. In its first stage, human symbols are grounded into neural networks. Later, infants learn to act in appropriate ways as symbols-for-a-child are grounded into culture. This second stage is further discussed below. Third, once a child's expressive powers develop, she will start to hear wordings: given brains and culture, symbols are grounded into first-person phenomenology.[4] Once wordings shape perception, they serve talk about language or, generally, speaking and acting *deliberately*. Over time, the results drive the functional reorganization of feelings, thoughts and actions. Being able to use wordings deliberately is crucial to rational action. As affect-based co-ordination is supplemented by *the said*, children master new routines (and games). Later, children use special modes of action to structure thoughts (Melser 2004). Much is learned by exploiting context to act epistemically (Cowley and MacDorman 2006). Interactivity enables bodies to use real time adjustments to discover 'organisational' constraints. Though neural predispositions influence ontogenesis, they function through concerted activity. Together, infants and caregivers orchestrate by sensitising to affect marked contingencies. They use co-action or, by definition, how one party used the context of another person's action to come up with something that could not have been done alone.[5] At 14 weeks a mother may be able to use, not touch, but the changing context of her body making her baby fall silent (Cowley et al. 2004). The baby attends to repeated action or, in Maturana's terms (1978), how each orients to the orienting of the other. As a result, circumstances are co-opted in strategic (joint) action. Learning to speak is initially separate from co-action. However, caregivers and infants use the rewards of interactivity to share use of contingencies. As infants manage adult displays, they draw on affect or, in Stuart's (2012) terms, how *enkinaesthesia* prompts us to orient to the felt co-presence of others. Later, these skills become enmeshed with those of vocalising.

In the first stage of human symbol grounding, brains rely on statistical learning. Before birth brains sensitise to rhythms of voices and languages. Infants show

[3]Here a symbol can be defined as a cultural constraint that serves in taking the measure or others and/or in controlling one's behaviour: many symbols are prosodic, gestural or enact what Goffman (1959) calls the interaction order.

[4]There is no clear evidence of when this occurs: however, there is abundant evidence that it is based on the skill of making and tracking phonetic gestures (Fowler and Rosenblum 1991; Fowler 2010). Further, since it is necessary to pretending it is likely that children begin to have the necessary experience in the second half of the second year.

[5]Wenger and Sparrow (2007) use experimental work to trace the social and bodily complexity of co-action—and its deep links with our *sense of agency*.

preferential response to the mother's voice (and face) and the *right kind of rhythm* (De Casper and Fifer 1980) and, remarkably, a story heard in the womb (De Casper and Spence 1986). While many animals discriminate, babies have skills in expressive co-ordination. Given rhythmic sensitivity, co-action soon falls under the baby's influence. This was first recognised in Bateson's (1971) work on the protoconversations that reveal 'intersubjectivity' (Trevarthen 1979). Context sensitive co-action is also stressed by Bråten (2007). More recently, the ability to co-ordinate expressive movements (including vocalisation) has been traced to grey matter in the brainstem which, before birth, controls morphogenesis (Trevarthen and Aitken 2001). As motivation develops, contingencies prompt a baby to use the rewards to interactivity to anticipate events. By three months, infants gain skills in controlling vocal, facial and manual expression. Norms already play a part in controlling their enkinaesthetic powers. Language and gesture (not to mention music and dance), thus share a neural basis (Willems and Hagoort 2007). As social actors, we rely on controlling expression in time: an infant uses affect to lock on to the movements by others and, by so doing, engages in dance-like co-action. Even congenitally blind infants move their hands to rhythmic patterns (Tønsberg and Hauge 1996). Those who hear normally, however, use its musical properties to discover the rewards of vocalising.

By three months events begin to show the influence of cultural symbols. Infants sensitise to signs of culture: using coordinated action human symbols are *grounded into culture*. Caregivers use infant gaze, smiles and other expressive gestures as indices of local norms that contribute to co-actional routines. Using both biological tricks and adult displays, infants gain a say in events. While infants and caregivers have *fun* together (see e.g., Stern 1971), affect allows interactivity to build contingencies into dyadic routines or *formats* (Bruner 1983). These help infants with when to initiate, what to expect and, of course, when to inhibit. Surprisingly, a three month old may 'do what its mother wants' by falling silent on command (Cowley et al. 2004); in an isiZulu speaking setting, the baby shows *ukuhlonipha* ('respect'). Dance-like interactivity helps the infant re-enact cultural values. This infant changes parental behaviour in ways that induce learning about situated events. Showing 'respect' (as we describe it), evokes a feeling tone. Without hearing words (or manipulating symbols), the baby comes to value *ukuhlonipha*. Given co-action, cultural contingencies connect with adult display. In this aspect of human symbol grounding, infant motivations exploit adult *experience*. Even if early normative behaviour uses biology (and neural systems that enable adults to shape infant expression) this is a developmental milestone. Before babies learn to reach for objects, caregivers will sometimes act as if their infants 'understand what is said'.

As symbols are grounded into culture, a 3–4 month is increasingly adjudged in terms of how well (or badly) she behaves. Given contingencies and rewards, she sensitises to *circumstances*. Instead of needing stimuli or cues, co-action changes how activity is modulated. For many reasons, the focus of development then switches to learning about objects. Late in the first year, however, the child discovers *secondary intersubjectivity* (Trevarthen and Hubley 1978) during Tomasello's (1999) *9 month revolution*. Bringing social and manipulative skills together, mediated or triadic behaviour emerges. Since language is co-ordinated activity, there is no need for the identification or recognition of inner intentions. Rather, it is sufficient that adults respond *as if* actions were representational. Infants use contingencies (and compressed information) by acting in ways that *seem* intentional. For example, Cowley (2007b) describes a mother who gets a nine-month old to fetch a block. Far from using inferences, the baby co-ordinates with maternal actions that include shifts of gaze, vocalising 'fetch' three times and using her whole body as a pointer. Fetching thus appears (and, perhaps, feels) intentional thanks to how the mother's vocalisation ('fetch') encourages norm-based activity. Further, if the child can mimic the sound (fetch), this opens up what Tomasello (1999) calls role reversal imitation. This is facilitated by independent concerns that include infant pleasure in self-produced vocalisations. As babbling shapes articulation, by 12 months, a baby 'repeats' syllables. Intrinsic motivations unite with skills and anticipated reactions. In the second year, a baby grasps 'facts' linking the normative with the physical. At times, she may do what is wanted. Once a toddler, she follows commands or, strikingly, directs adult attention and actions. She grasps and utters (what adults hear as) 'words'. As wordings fuse with first-order activity, human agency emerges.

As an infant begins to walk, she is becoming to adopt social roles. While far from reasoning, she draws on virtual patterns and social norms. Given a simple toy, a 12 to 18 month old will enact cultural expectations. In an unpublished study, a French and an Icelandic child-mother dyad played the 'same' game together over several weeks (Sigtryggsdóttir 2007). Whereas the French dyad used this in having fun, the Icelandic partners treated it as goal directed activity. Each baby learned how to elicit rewards. Strikingly, when the Icelandic mother failed to participate, the baby would sometimes self-applaud. Plainly, she exploited—not just affect—but (shared) goals. She has become attuned to the values of her world. Co-ordination enables both parties to use maternal displays of cultural values to organise activity. By participating in routines based in local ways of life, a child learns about fun as well as rationality. Far from relying on sound-patterns, the baby uses rewards that co-vary with what is intended (e.g., 'show respect!'). Quite unknowingly, the child orients to other-directed functions of caregiver verbal patterns. Yet, no one year old *hears* wordings. It is only later that utterances come to be heard as utterances *of* patterns. Early on, first-person experience arises in dynamics of co-action (Cowley 2011a, 2011b). Time passes before a child discovers the potential advantages of using wordings as if they were things.

Coming to *hear* verbal patterns changes human (inter) action and perception. While its neural basis lies in tracking phonetic gestures that feature rich variability, adults use perceptual images as 'words'. All of us have some experience of what is called 'private' thinking by means of public symbols. The skill appears in Piaget's *symbolic stage* when, not by coincidence, infants discover pretending. By hearing more than sound, they discover a 'magical' aspect of language. A two year may say (to a banana), *"hello, who's there?"* Without being able to *hear* telephone talk (a remarkable cognitive skill), such pretending could not arise. Given this perceptual learning, a child learns both to get others to do what *she* wants and to use self-directed speech to shape her action. Whereas language sensitive bonobos can follow novel instructions like 'go and put the orange in the swimming pool' (Savage-Rumbaugh et al. 1998), children excel in different skills. Given their biases, they use wordings as social resources. Thus, unlike bonobos, humans share attention for its own sake. By age 3 a human child will not only follow instructions but will language differently with peers and in pre-school. She will choose when not to conform: for a child wordings are social affordances. This is just as important in the life of a social actor.

Just as children are not born talking, they are not born rational. Rather, the skills that shape language and reason arise as we identify with aspects of what we hear ourselves say. Co-action prompts infants to orient to local modes of expression. Given a developmental history, layers of agency accumulate together with a history of decision-making. As we redeploy neural resources, we draw on biologically guided interactivity. We learn from a history of anticipating how others will use norms. To act as sophisticated agents training must hone our biological capacities. Human symbol grounding makes us into norm-recognising agents not unlike symbol processors. This depends on individually variable skills in using the language stance to manipulate wordings. Indeed, without this combination, communities would be unable to identify themselves as speakers of a specific language. Without being able to describe words (and rules) as entities that pertain to an autonomous system (e.g. English): we would not believe in abstractions like *minds, selves or societies*.

Using the Resources of Reason

Public language permits *objectively valid* judgements. For Craik (1943), this is exemplified by when, in bridge-building, language is put to symbolic *use*. How this is conceptualised is, of course, a theoretical matter. Since Craik ascribed this to the brain, he viewed language itself as representational. However, the distributed view offers a parsimonious alternative. People learn to refer: they depend on connecting talk about language with languaging (i.e., activity). We learn to take the perspective of the other while linking articulatory patterns with items of shared attention. Once we begin to take a language stance, we hear wordings as wordings that serve to pick out things as things. With practice, we learn to refer or, in other terms, how languaging can be used to pick out objects, persons, events etc.

The skills of a rational human agent depend on both real-time coordination and using the language stance to exploit cultural resources. Given the phenomenological status of wordings, they can be used both literally and in fun. This is because, since they arise from interactivity, they are integrated with action and expression as we enact relationships. In contrast to the fixity of computational symbols, wordings gain effectiveness from flexibility and vagueness. Sense-making arises as they are jointly crafted as persons orient to circumstances and each other. Unlike symbols used in computers, they bear on what people are doing. As part of language flow, interactivity or, in lay terms, *how* we speak is meaningful. Rightly, therefore, many contrast language with man-made codes. Unlike Morse, for example, language is neither monological, disembodied nor dependent on design. Unlike man made codes, language is dialogical (Linell 2005, 2009). Further, given its embodiment, brains ensure that language is enmeshed with action (e.g., Willems and Hagoort 2007). As one thread in coordination, its literal or denotational meaning is often marginal (Maturana 1978). Circumstances add to how, as living bodies coordinate, we make and construe linguistic signs (Love 2004). Dynamics make language irreducible to words, utterance-types, usage-patterns and so on (Harris 1981; Love 2004). As argued here, people—not language—exploit acts of reference: representationalism is not necessary to cognition (e.g., Anderson 2003; Thompson 2007; Stewart et al. 2010). Indeed, computing systems face the *symbol grounding* problem (Harnad 1990): computations are meaningless to machines. Worse still, where grounded by design, symbols fail to pick out facts. No (currently imaginable) robot could 'know' that, 'Put out the light' is irrelevant to, say, what is in the fridge or a US president's concerns (see, MacDorman 2007). They face *the frame problem* and, for this reason, robots are increasingly designed in ways that enable them to use people to connect with the world.[6]

In contrast to symbol processing, public and collective behaviour enacts skilled co-ordination. Even infants use actions and utterances as representations (e.g., in pointing). During co-action, adults treat infant doings as intentional. Where the infant identifies relevant features, repetition shapes conventional behaviour. For the adult, the infant acts representationally. To extend its cognitive powers, therefore, the baby tracks contingencies. Indeed, as the cultural world increasingly aligns to joint behaviour, the baby learns about co-action. Later, infants come to *hear* words by using interactivity to track contextual indices of local norms. Further, this is a likely basis for using wordings representationally. By using a language stance they become thing-like entities that sustain *belief* in virtual constructs (Cowley 2011a, 2011b). Accordingly, they can be used both as expected and in transgressive ways. Indeed, both approaches lead to sense-making because of how the results are integrated with coordination. Once we perceive second-order constructs as thing-like, we can generate *thoughts* by modulating how we speak or use a pencil to make pictures. Although partly verbal (and symbolic), interactivity connects bodies, activity and cultural experience. Unlike Morse (or computer programs), language is dialogical, multi-modal and realises values (Hodges 2007). Eventually, wording-based reality links with the resources of an interpretative hermeneutic community.

The spread of language prompts social actors to reproduce society. Within a cultural space, children use interactivity to grasp how symbols serve social action. By coming to anticipate how others speak, they discover Wittgenstein's *agreements in judgement* (1958: §242). Using coordination, they develop social strategies that depend on connecting words, circumstances, and the music of expression. In this way, unmeasurable virtual patterns take on cognitive power. Like other living things, we depend on compressed (Shannon) information. In Dennett's (1991) terms, humans

---

[6]One of the most remarkable facts about robots is that, already, they use human consciousness: this is exemplified, for example, when they learn to discriminate what we—but not they—see as colours (see, Cowley 2013). Building on this view, it can be argued that robots are of importance as linguistic informants (Cowley 2008).

use *real-patterns* that include not only cases like gravity and colours but also wordings. Because of phenomenological status, we can use for example, as ensembles of norms that reflect on other people's expectations. Davidson's (1997) view of the role of thought and language is similar. He proposes a uniquely human *framework* (p. 27): "The primitive triangle, constituted by two (and typically more than two) creatures reacting in concert to features of the world and to each other's reactions, [...] provides the framework in which thought and language can evolve. Neither thought nor language, according to this account, can come first, for each requires the other".

Others concur that thought and language co-emerge from interactions. For Maturana (1978), an agent's sense of self fuses with verbal patterns: *structural coupling* allows new-borns to engage with caregivers. Their languaging soon becomes oriented to *types* of circumstance (and thus a *consensual domain*). This generates (observer dependent) opportunities for sense-making. Gradually, however, perceiving, feeling and acting are integrated with normative aspects of language. As neural functions change, individuals become *speakers*. In our terms, experience allows discrepant awareness to shape skills based on taking the language stance. A child's sense of self uses coordinated action to link cultural resources with individual skills. For example, we talk about talk, develop narratives, and make up autobiographical memory. By using discrepant awareness, we link circumstances with the past, the possible and the future. It is not the brain but, rather, languaging that underpins reference. Even bridge-building integrates symbolic, practical and skill-based knowledge based on a life history of games that make us (more) rational. Using standardisation, dictionary writing and education, (increasing) weight falls on literal meaning. As this becomes familiar, the language stance favours a detached 'point of view' and more body-centred control of thoughts, feelings and actions—provided that we reproduce social 'reality'.

Human Agency Naturalised

In naturalising human agency, we claim that experience of co-ordinating shapes our cognitive, social, and linguistic skills. Thereby we reformulate Durkheim's old claim that the social explains the social, namely by explaining how biological members of sapiens develop the dimensions of sociological agency. While bodies are pre-adapted for cultural learning, interactivity prompts brains to compress information by orienting to verbal patterns, artifacts and norms. We refer by calling up the past, the possible and the future. This is, of course, dependent on institutions and artifacts. Social relations thus underpin reasoning and, of course, skills in making what count as objectively valid or wise judgements. Though we use the results to model social phenomena, their basis lies beyond the brain. We depend on coordinating spontaneously while making judicious use of the language stance and the resources of reason. Though languaging retains its importance to face-to-face thinking, in many other settings, weight falls on treating wordings as wordings. As Piaget

(1962) shows, we come to grasp games of marbles or, later, take part in literacy practices. We increasingly use the language stance to participate in the assemblages that enact joint projects. Human agency is partly eusocial. Its develops from a kind of *fission*: as biological infants become persons, a chain of interactivity transforms what they can do. As this happens, they increasingly discern uses for cultural resources that serve in both individual and collective endeavours.

The transformatory power became especially clear when a bonobo chimpanzee, Kanzi, was raised in a human-like environment (see, Savage-Rumbaugh et al. 1998). Not only did he gain from computer access to verbal resources but these were coupled by close attention from human carers. Bonobo symbol grounding made Kanzi strikingly (partly) human-like.[7] The case contrasts with Davis's (1949) description of Anna, whose first years of life lacked social embedding and emotional care. She did not speak, could not eat on her own and never laughed or cried. Lack of human company deprived her of opportunities for learning from how people use co-action in orienting to social norms. She never used interactivity in feeling out a cultural world and, as a result, failed to develop the cognitive powers used in social life. Unlike a normal human actor (or Kanzi), her actions were loosely constrained by culture. In short, sociological agency arises as language becomes a dimension of the person. Eighteenth century tradition wrongly plucked words from the world. Language is no transparent medium because, contra Pinker, wordings are not located in the mind (or brain). Rather, they are part of public activity between people, activity that allows even a 14 week old to use co-ordination to show ukuhlonipha ('respect'). The baby does not 'encode' meanings or propositions but, rather, learns from the routines of everyday co-action.

A Sketch of Social Fission

For the social sciences, interactivity and languaging are conceptually important. Although everyday language may be the necessary basis for modelling macro-social phenomena, it seems inappropriate to the micro-social domain. The models of social actor theory (Boudon 1981; Coleman 1990; Hedström 2005), like those of code linguistics and the computational theory of mind, ignore the world of embodied, conscious beings. In appealing to social fission, we thus naturalize how the social grounds the social. Rather than treat genes and brains as the origins of reason, we argue that children use interactivity to develop locally appropriate kinds of agency. They draw on experience and, crucially, use the language stance to grasp how people, circumstances and situations vary. Brains and genes predispose infants for cultural learning that, by hypothesis, depends on compressing (Shannon) information. They learn about social (and other) affordances as coordination produces experience with norms, artifacts and wordings. Indeed, the flaws in individual rationality speak

[7]This depends on the observer's point of view: in many respects, Kanzi remains distinctly a bonobo.

strongly *against* ontological individualism. Rationality derives from social relations: it is a feature of the cultural and institutional environment that drives biological humans to make imperfect use of (what count as) objectively valid judgements.

Since symbolic models capture macro-social patterns, biological humans *discover* the resources of reason. Our agency is made and not born; it emerges from both the physical world and affordances such as languages, artifacts and social institutions. Far from centring on a body (or brain), it depends heavily on how languaging enacts social relations. Though, often, we cannot be literal, judicious use of the language stance brings rich rewards. Combined with appraisal and interactivity, we unearth the value of cultural resources. While sometimes acting individually, joint projects tend to dominate our lives: the artificial matters greatly to human agency. It is thus to be expected that coordination serves to make strategic plans. Following Darwinian logic, it is not at all surprising that social affordances are selected as a result of enacting social relations. This may be why most cultures develop, for example, ways of displaying and recognising kinds of *trust* and *reciprocity*. Interactivity gives rise to a selection history that links up languages, institutions and social norms. Human agents develop intuitive or expert skills alongside those based on the resources of reason. Since human *nature* is so flexible, it is an error to use 'Hobbes's Problem' as evidence for the difficulty of coordination. That said, our limited rationality does create practical problems of aggregation (Spiekermann 2013; Ben-Naim et al. 2013). Rather than view this as a symptom of inherent selfishness, it shows that humans need complex resources that provide results as they move in and out of social aggregates. These make human cognition partly eusocial—much depends on collective modes of action that link the artificial domains of languages, artifacts and institutions.

Interactivity in Human Agency

By acknowledging that cognition cannot be explained by processes within the brain, we move towards a new sociocognitive science. Human agency is constantly re-enacted as interactivity links us with the world. As we do so, we move in and out of social aggregates that draw on languages, artifacts and social institutions. We find our way through the wilds, talk and, for that matter, use computers and develop skills in flying planes. Human agency is not to be identified with the agent. Since it derives from a history of engagement with the world, agency can be traced to four sets of constraints. First, as physico-chemical systems, we exert (and suffer) physico-chemical effects. Second, as living beings, the boundary conditions of our lineage shape the parameters that result from growth, action, learning and development. Third, as human agents, we develop biophysical skills that exploit artificial constraints associated, above all, with artifacts, languages and institutions. Rather than function as boundary conditions, these flexible resources allow us to pursue individual and collective endeavours. Finally, as living subjects, we make and construe artificial affordances. Thus, we are not social actors, languages are not codes and minds are not symbol processors. In rejecting all such organism-centred views, the distributed perspective holds out the prospect of reintegrating biosemiotics, cognitive psychology, linguistics and the social sciences. The core idea is that our becoming can be traced to interactivity that links agents in larger aggregates within a common world. Although creativity gives rise to artifacts, inscriptions and public performances, its basis lies in how biosocial agents mesh temporal scales while using interactivity. Remarkably, it seems that a single sociocognitive system enables brains, languages and societies to conspire in prompting human bodies to make partial sense of the world. This is crucial to the goals of the field. On the one hand, as noted above, we need to clarify how people come to hear and exploit wordings. On the other, this opens up the much broader question of how phenomenological experience links with organisation in other time-scales. In short, what can be described in language must be traced, on the one hand, to the rapid time-scales of interactivity and neural processes and, on the other, the slow scales that allow groups to differentiate in ways that drive cultural selection. It is there that fission prompts individuals to become the persons that we are. Just as slow scales constrain faster ones, the rapid processes of interactivity and languaging engender human agency—agents who ceaselessly re-evoke the past to explore the adjacent possible.

Acknowledgements Martin Neumann's contribution was undertaken within the Project 'Emergence in the Loop' (EmiL: IST-033841) funded by the Future and Emerging Technologies programme of the European Commission, in the framework of the initiative "Simulating Emergent Properties in Complex Systems". Stephen Cowley's draws much on the opportunity to contribute to a programme of scientific workshops on Agency and languaging (2010–2012) that were organised by the University of Jyväskylä (and sponsored by the Finnish Ministry of Culture).

References

Anderson, M. L. (2003). Embodied cognition. *Artificial Intelligence*, *149*, 91–130.

Bateson, M. C. (1971). The interpersonal context of infant vocalization. *Quarterly Progress Report of the Research Laboratory of Electronics*, *100*, 170–176.

Ben-Naim, J., Bonnefon, J. F., Herzig, A., Leblois, S., & Lorini, E. (2013). Computer-mediated trust in self-interested expert recommendation. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Bickhard, M. (2007). Language as an interaction system. *New Ideas in Psychology*, *25*, 171–187.

Boudon, R. (1981). *The logic of social action: an introduction to sociological analysis*. London: Routledge.

Bråten, S. (2007). Altercentric infants and adults: on the origins and manifestations of participant perception of others' acts and utterances. In S. Bråten (Ed.), *On being moved: from mirror neurons to empathy* (pp. 111–135). Amsterdam: Benjamins.

Bruner, J. (1983). *Child's talk: learning to use language*. New York: WW Norton Press.

Coleman, J. (1990). *Foundations of social theory*. Cambridge: Belknap.

Cowley, S. J. (2007a). How human infants deal with symbol grounding. *Interaction Studies*, *8*, 81–104.

Cowley, S. J. (2007b). The cradle of language: making sense of bodily connections. In D. Moyal-Sharrock (Ed.), *Perspicuous presentations: essays on Wittgenstein's philosophy of psychology* (pp. 278–298). Houndmills: Palgrave Macmillan.

Cowley, S. J. (2007c). Cognitive dynamics and distributed language. *Language Sciences*, *29*, 575–583.

Cowley, S. J. (2008). Robots: the new linguistic informants? *Connection Science*, *20*, 359–369.

Cowley, S. J. (2009). Language flow: opening the subject. *Cognitive Semiotics*, *4*, 64–92.

Cowley, S. J. (2011a). *Distributed language*. Amsterdam: Benjamins.

Cowley, S. J. (2011b). Taking a language stance. *Ecological Psychology*, *23*, 1–25.

Cowley, S. J. (2013). Naturalizing language: human appraisal and technology. *AI & Society*. doi:10.1007/s00146-013-0445-3.

Cowley, S. J., & MacDorman, K. F. (2006). What baboons, babies and Tetris players tell us about interaction: a biosocial view of norm-based social learning. *Connection Science*, *18*, 363–378.

Cowley, S. J., Moodley, S., & Fiori-Cowley, A. (2004). Grounding signs of culture: primary intersubjectivity in social semiosis. *Mind, Culture and Activity*, *11*, 109–132.

Craik, K. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.

Cronin, B. (2004). Bowling alone together: academic writing as distributed cognition. *Journal of the American Society for Information Science and Technology*, *55*, 557–560.

Davis, K. (1949). *Human society*. New York: Macmillan.

Davidson, D. (1997). Seeing through language. *Royal Institute of Philosophy Supplement*, *42*, 15–28.

Dennett, D. C. (1991). Real patterns. *The Journal of Philosophy*, *88*, 27–51.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mother's voices. *Science*, *208*, 1174–1176.

DeCasper, A. J., & Spence, M. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development*, *9*, 133–150.

Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4*, 429–452.

Durkheim, E. (1895 [1998]). *Rules of sociological method*. Basingston: Macmillan.

Fowler, C. A. (2010). Embodied, embedded language use. *Ecological Psychology*, *22*, 286–303.

Fowler, C. A., & Rosenblum, L. D. (1991). The perception of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 33–59). Hillsdale: Erlbaum.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Goffman, E. (1959). *The presentation of self in everyday life*. New York: Doubleday Anchor.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, *42*, 335–346.

Harris, R. (1981). *The language myth*. Oxford: Duckworth.

Hedström, P. (2005). *Dissecting the social*. Cambridge: Cambridge University Press.

Hodges, B. (2007). Good prospects: ecological and social perspectives on conforming, creating and caring in conversation. *Language Sciences*, *29*, 584–604.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.

Järvilehto, T. (1998). The theory of the organism-environment system: I. Description of the theory. *Integrative Physiological and Behavioral Science*, *33*, 321–334.

Järvilehto, T., Nurkkala, V. M., & Koskela, K. (2009). The role of anticipation in reading. *Pragmatics & Cognition*, *17*, 509–526.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*, 513–549.

Kravchenko, A. (2007). Essential properties of language: why language is not a digital code. *Language Sciences*, *29*, 621–650.

Linell, P. (2005). *The written language bias in linguistics: its nature, origins and transformations*. London: Routledge.

Linell, P. (2009). *Rethinking language mind and world dialogically: interactional and contextual theories of human sense-making*. Charlotte: Information Age.

Locke, J. (1690 [1975]) *An essay concerning human understanding*. Oxford: Oxford University Press. Edited by P. H. Niddich.

Love, N. (2004). Cognition and the language myth. *Language Sciences*, *26*, 525–544.

MacDorman, K. (2007). Life after the symbol system metaphor. *Interaction Studies*, *18*, 143–158.

Maturana, H. R. (1978). Biology of language: the epistemology of reality. In G. A. Miller & E. Lenneberg (Eds.), *Psychology and biology of language and thought: essays in honor of Eric Lenneberg* (pp. 27–63). New York: Academic Press.

Melser, D. (2004). *The act of thinking*. Cambridge: MIT Press.

Pattee, H. H. (1969). How does a molecule become a message? *Developmental Biology. Supplement*, *3*, 1–16.

Pattee, H. H. (1997). The physics of symbols and the evolution of semiotic controls. In *Proceedings from the workshop on control mechanisms for complex systems: issues of measurement and semiotic analysis*, Santa Fe Institute Studies in the Sciences of Complexity. Redwood City: Addison-Wesley.

Piaget, J. (1962). *Play, dreams and imitation in childhood*. New York: Norton.

Port, R. (2010). Rich memory and distributed phonology. *Language Sciences*, *32*, 43–55.

Rączaszek-Leonardi, J. (2009). Symbols as constraints: the structuring role of dynamics and self-organization in natural language. *Pragmatics & Cognition*, *17*, 653–676.

Raguso, R., & Willis, M. (2002). Synergy between visual and olfactory cues in nectar feeding by naïve hawkmoths Manduca Sexta. *Animal Behavior*, *64*, 685–695.

Reddy, M. J. (1979). The conduit metaphor: a case of frame conflict in our language about language. In A. Ortony (Ed.), *Metaphor and thought* (pp. 284–297). Cambridge: Cambridge University Press.

Ross, D. (2007). H. sapiens as ecologically special: what does language contribute? *Language Sciences*, *16*, 710–731.

Savage-Rumbaugh, S. E., Shanker, S. G., & Taylor, T. J. (1998). *Apes, language and the human mind*. Oxford: Oxford University Press.

Sigtryggsdóttir, T. (2007). *Guiding attention leads to changes in child's manifest practical understanding*. Unpublished dissertation, University of Hertfordshire.

Spiekermann, K. (2013). Judgement aggregation and distributed thinking. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Steffensen, S. V. (2011). Beyond mind: an extended ecology of languaging. In S. J. Cowley (Ed.), *Distributed language* (pp. 185–210). Amsterdam: Benjamins.

Stern, D. N. (1971). A micro-analysis of mother-infant interaction: behaviors regulating social contact between a mother and her three-and-a-half-month-old twins. *Journal of American Academy of Child Psychiatry*, *10*, 501–517.

Stewart J., Gapenne, O., & Di Paolo, E. (2010). *Enaction: toward a new paradigm for cognitive science*. Cambridge: MIT Press.

Stuart, S. A. J. (2012). Enkinaesthesia: the essential sensuous background for co-agency. In Z. Radman (Ed.), *The background: knowing without thinking* (pp. 167–186). Houndmills: Palgrave Macmillan.

Thibault, P. J. (2011). First-order languaging dynamics and second-order language: the distributed language view. *Ecological Psychology*, *23*, 210–245.

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

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Cambridge University Press.

Tønsberg, G. H., & Hauge, T. S. (1996). The musical nature of prelinguistic interaction. The temporal structure and organization in co-created interaction with congenital deaf-blinds. *Nordic Journal of Music Therapy*, *5*, 63–75.

Trevarthen, C. (1979). Communication and co-operation in early infancy: a description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech* (pp. 321–347). Cambridge: Cambridge University Press.

Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity: research, theory and clinical applications. *Journal of Child Psychology and Psychiatry*, *42*, 3–48.

Trevarthen, C., & Hubley, P. (1978). Secondary intersubjectivity: confidence, confiding and acts of meaning in the first year. In A. Lock (Ed.), *Action, gesture, and symbol* (pp. 183–229). New York: Academic Press.

Wenger, D., & Sparrow, B. (2007). The puzzle of coaction. In D. Ross, D. Spurrett, H. Kinkaid, & L. G. Stephens (Eds.), *Distributed cognition and the will* (pp. 17–41). Cambridge: MIT Press.

Willems, R., & Hagoort, P. (2007). Neural evidence and the interplay between language, gesture and action: a review. *Brain and Language*, *101*, 278–289.

Wittgenstein, L. W. (1958). *Philosophical investigations* (2nd ed.). Oxford: Blackwell.

Chapter 3
Judgement Aggregation and Distributed Thinking

Kai Spiekermann

Abstract In recent years, judgement aggregation has emerged as an important area of social choice theory. Judgement aggregation is concerned with aggregating sets of individual judgements over logically connected propositions into a set of collective judgements. It has been shown that even seemingly weak conditions on the aggregation function render it impossible to find functions that produce rational collective judgements from all possible rational individual judgements. This implies that the step from individual judgements to collective judgements requires trade-offs between different desiderata, such as universal domain, rationality, epistemological quality, and unbiasedness. These dilemmas challenge us to decide which conditions we should relax. The typical application for judgement aggregation is the problem of group decision making, with juries and expert committees as stock examples. However, the relevance of judgement aggregation goes beyond these cases. In this survey I review some core results in the field of judgement aggregation and social epistemology and discuss their implications for the analysis of distributed thinking.

Thinking is often taken as an activity exercised by individuals. In recent years, however, it has been acknowledged that thinking can also be a collective process. It is not only individuals who process information, take stances, and make decisions—groups can do this, too. For instance, a court jury needs to gather information, reach collective stances on the information available, and make a decision on the sentence. The same is true for cabinets, expert panels, shipping crews, air-traffic controllers, appointment committees, etc. Individuals can differ in their ability to process information rationally and reach correct decisions. Similarly, groups can differ in their success to arrive at correct decisions, and they may arrive at these decisions in a rational or in an irrational way. In this sense groups are engaged in collective thinking.

It is difficult to observe how individuals process complex information and arrive at a decision. For groups, however, this process is more transparent. Psychologists and social scientists can observe how groups deliberate, how they form judgements,

K. Spiekermann (✉)
Department of Government, London School of Economics, Houghton Street, London, WC2A 2AE, UK
e-mail: k.spiekermann@lse.ac.uk

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_3, © Springer-Verlag London 2013

and how they finally reach decisions. These collective decision processes can be compared and evaluated. Some processes are obviously epistemically poor and irrational: For instance, a court jury should not throw a coin to decide whether a defendant is guilty; an expert panel should (arguably) not randomly select one expert to make all decisions, a cabinet should not always choose the course of action with the least support, etc. But it is more difficult to determine good collective decision procedures, and we will see that it is often impossible to determine collective decision procedures that meet some seemingly harmless desiderata.

Different strands of literature have discussed the nature of distributed thinking. One strand, inspired by the psychologist Lev Vygotsky (1962, 1978), analyses distributed thinking in relation to distributed cognition. In recent years, this research paradigm was advanced by Edwin Hutchins (1995) and his influential study of "cognition in the wild". Research on distributed cognition was also influenced by Clark and Chalmers's concept of the "extended mind" (Clark and Chalmers 1998; Clark 1997). This approach emphasizes that the boundaries between mind and world are often difficult to draw, and that organisms reshape their environment to solve the problems they encounter. The mind is a "leaky organ", as Clark (1997, p. 53) puts it, "mingling shamelessly with body and with world".

Another strand of literature that tackles the phenomenon of distributed thinking draws on concepts from social choice theory and social epistemology. Goldman (2004) and List (2005, 2012) have pointed out that there are at least two dimensions on which a group's performance as a thinking system can be measured. First, the group can succeed or fail to be rational, where rationality is understood as avoiding logical contradictions in the judgements the group makes. Second, the group can be more or less successful in reaching correct decisions, given the information available. The first dimension poses a "rationality" or "coherence" challenge, the second a "knowledge" or "correspondence" challenge to the group. The rationality challenge can be explored with tools provided by social choice theory, the knowledge challenge with generalizations of the Condorcet Jury Theorem and the information pooling literature.

Social choice theory systematically investigates the processes for the aggregation of individual information into collective information. The classical problem for social choice theory is the aggregation of preferences. The famous Arrow Theorem shows that there is no aggregation procedure to map the individual preferences to collective preferences that meets some seemingly harmless and arguably normatively desirable conditions (Arrow 1963). While the aggregation of preferences is of great importance for welfare economics, it is not quite the right framework to address distributed thinking. However, recently the social choice framework has been extended to the more general question of judgement aggregation. Judgement aggregation investigates different procedures to aggregate individual judgements to collective judgements. Again, impossibility results arise, posing challenges for distributed thinking.

Related to the field of judgement aggregation are considerations regarding the epistemic quality of different aggregation procedures. The discussion starts with Condorcet's famous observation that large groups tend to make correct dichotomous choices. But Condorcet's ideas can be extended to other choice situations as well. If information is distributed between agents, and these agents need to arrive at a joint decision based on the information, then one can ask which procedures are best suited to aggregate this information to maximize the probability of a correct decision.

Returning to the two strands of literature mentioned above, it appears that the distributed cognition approach on the one hand, and the social choice and epistemology approach on the other, talk past each other. So far, the exchange between the two approaches has been limited. This chapter does not attempt a reconcilliation or propose a new synthesis. The rather more modest goal is to introduce some basic ideas from the field of judgement aggregation and consider the upshot for distributed thinking. The chapter is in 4 sections. I start by explaining the problem posed by the "discursive dilemma" and how it pertains to distributed thinking. The section *Impossibility Results and Escape Routes* generalizes by approaching problems of judgement aggregation more formally. In the section *An Epistemic Perspective* I analyse the epistemic performance of different judgement aggregation procedures. I discuss the relation between rationality, consistency, and distributed thinking in the section *Distributed and Consistent Thinking*. At this point, I will return to the relation between distributed cognition and social choice theory, and discuss how these two approaches may relate. More specifically, I will argue that judgement aggregation provides a framework for the analysis of distributed thinking, despite charges that it is too reductionist to be of interest.

The Discursive Dilemma

A central problem that has triggered much work in the field of judgement aggregation is the so-called "doctrinal paradox" (Kornhauser and Sager 1986) or, more generally, the "discursive dilemma" (List and Pettit 2002). I start by describing two examples that illustrate the problem.

Consider three MI5 officers who have to evaluate whether an observed suspect is planning to build a bomb. The three officers assess the situations by forming judgements on the correctness of three propositions:

- The suspect has bought fertilizer (P).
- If the suspect has bought fertilizer, it follows that the subject plans to assemble a bomb ($P \rightarrow Q$).[1]
- The suspect plans to assemble a bomb (Q).

These propositions are logically connected. For instance, if an officer believes that the subject has bought fertilizer, and if she also believes that if the subject has bought fertilizer then the subject is building a bomb, then the officer must also hold that the subject plans to build a bomb. If she does not, the officer's judgements would be inconsistent.

[1] For the example discussed here we can take $\rightarrow$ as the material conditional.

Table 3.1 An example of the discursive dilemma

Officer	P	$P \rightarrow Q$	Q
1	true	true	true
2	true	false	false
3	false	true	false
Majority	*true*	*true*	*false*

We assume that all officers (individually) hold consistent sets of beliefs, i.e. they do not contradict themselves, and that they make judgements on all propositions at stake. One possible constellation of consistent individual judgements over these three propositions is shown in Table 3.1.

Officer 1 thinks that the suspect has bought fertilizer, that if the subject has bought fertilizer he is planning to build a bomb, and consequently thinks the suspect builds a bomb. Officer 2 also thinks that the suspect has bought fertilizer, but disagrees with the claim that buying fertilizer implies that the suspect builds a bomb, and thinks that the suspect does not build a bomb. Officer 3 disagrees with his two colleagues about whether the suspect has bought fertilizer. He believes that if the suspect had bought fertilizer he would be building a bomb. But since he has not, officer 3 can hold (for whatever reason, as no conclusion follows from the premises) that the suspect does not build a bomb.

The problem in this situation is that the three officers will find it difficult to determine their joint stance as an investigative unit. A majority vote on each proposition yields the results as stated in bottom row of Table 3.1. A majority thinks that the suspect has bought fertilizer, a majority thinks that if the suspect has bought fertilizer he is assembling a bomb, but a majority also thinks that the suspect is not building a bomb. Thus the majority judgements are contradictory. This contradiction instantiates one version of the discursive dilemma.

Consider a second example to demonstrate that the discursive dilemma comes in different forms. Here a team of detectives has to decide whether to bring charges forward against a suspected murderer. The three detectives consider the following propositions:

- The murder weapon is identified (P).
- The suspect had a motive (Q).
- The suspect should be charged (R).
- Charges should be brought forward if and only if the weapon is identified and the suspect had a motive ($P \wedge Q \leftrightarrow R$).

We assume that the three detectives all agree on the last proposition, which one can interpret as a universally agreed doctrine. They disagree, however, on the other three propositions, as Table 3.2 shows.

As in the first example, each individual position is consistent, but the majority position is not. Holding P, Q, and $P \wedge Q \leftrightarrow R$ to be true, but R to be false, is a contradiction. The question is: how should the three detectives agree on a joint position?

Table 3.2 The discursive dilemma in conjunctive form

Detective	P	Q	$P \wedge Q \leftrightarrow R$	R
1	true	true	true	true
2	true	false	true	false
3	false	true	true	false
Majority	*true*	*true*	*true*	*false*

In these examples we can recognize some features of the discursive dilemma and problems of judgement aggregation more generally. First, the dilemmas described here are fairly realistic in the sense that there are many situations in which groups of people hold judgements over different logically connected propositions and have to form a joint position on these propositions. Second, the examples can easily be extended to groups with more than 3 agents. Third, the individual judgements the agents hold are not unusual or unreasonable. Fourth, the dilemma only arises for certain judgement profiles, but the possibility of their occurrence challenges us to find judgement aggregation procedures that can deal with these situations.

How does the discursive dilemma pertain to distributed thinking? A thinking system understood in a minimal way is a system that takes inputs and produces outputs by processing these inputs. A distributed thinking system can be understood as a group of thinkers who coordinate their thinking activities. Since the thinking is distributed, one can expect every single thinker to do some thinking on their own. However, to function as a thinking system it is necessary to aggregate the information available to the single thinkers and produce a collective output. Judgement aggregation is a model of such a process: Each individual is a single thinker with stances on certain propositions. Since the single thinkers are part of a distributed thinking system, the system must aggregate their stances on the propositions and produce a collective stance. In the same way as we want single thinkers to be rational, we also require a system of distributed thinking to be rational. Judgement aggregation maps out the logical space of possible aggregation procedures and informs us of the options and constraints for distributed thinking.

The notion of "thinking" in the analysis offered here is deliberately minimal. It presupposes only that thinkers assign truth values to each proposition and that thinkers correctly apply propositional logic. In addition, the distributed thinking system must be able follow an aggregation rule. The problems arising from this simplified notion of thinking are neither trivial nor simple, and it is worthwhile to start with simple examples before moving on to more complex analyses. This minimal notion of thinking deliberately omits many other aspects of human thinkers: First, people can have degrees of beliefs, not just dichotomous judgements. Second, thinking does not only involve beliefs, but also desires. Third, a complete picture of human thinking would also incorporate intentions, emotions, and consciousness. Nonetheless, I argue that richer notions of thinking can be set aside for now. They can be set aside because even the minimal notion of thinking used in this chapter raises interesting questions about the rationality and epistemic quality of distributed thinking systems.

Impossibility Results and Escape Routes

I now describe the problem of judgement aggregation more generally and explain List and Pettit's (2002) impossibility result. Each individual has a *set of judgements* on a given *agenda*. The agenda contains all propositions in question and their respective negations. For the impossibility result to arise, the agenda must be sufficiently complex, that is it must contain at least two propositions P and Q and either $P \wedge Q$, $P \vee Q$, or $P \rightarrow Q$ (and their negations). An individual set of judgements must be complete (so that for all items on the agenda, it contains either the proposition or its negation), it must be consistent and deductively closed. If these three conditions are met, we call a judgement set fully rational. All the individual sets of judgements together form a judgement profile. For instance, Tables 3.1 and 3.2 state specific judgement profiles.

The aim of judgement aggregation is to proceed from judgement profiles to a *collective judgement set*. We assume that the collective judgement set must also be consistent, complete, and deductively closed (this is called the *collective rationality condition*, see e.g. List 2012) and that the aggregation function never fails to produce output. An *aggregation function* has all possible judgement profiles as domain and all possible collective sets of judgements as co-domain, that is, it maps judgement profiles onto collective sets of judgements. Put differently: An aggregation function takes a judgement profile as input and gives one fully rational collective set of judgements as output. List and Pettit describe three desiderata that an aggregation function should meet:

Universal Domain. The aggregation function accepts as input all logically possible judgement profiles, as long as all individual judgement sets are consistent, complete, and deductively closed.

Anonymity. The aggregation function is not responsive to permutations of judgement sets in the profile. This means that the outcome should not change if we shuffle the agents, but leave everything else unchanged.

Systematicity. The result of the aggregation function for any proposition depends only on the judgements made on this proposition, and the pattern of dependence is the same for all propositions.

Universal Domain is an immediately convincing desideratum: The aggregation function should be able to aggregate all logically possible profiles, as long as all individuals hold fully rational judgements. If the aggregation function did not have a universal domain it would fail to aggregate some judgement profiles that can occur, and there is no good reason to rule out any judgement profiles ex ante.

Anonymity is also a rather convincing desideratum for many aggregation problems. The intuitive appeal behind anonymity is that it ensures the equal treatment of all judgement sets, no matter who holds them. For example, anonymity rules out that the aggregation function always follows the judgement set of one individual, that is it rules out 'aggregation' by letting one agent be the dictator.

The systematicity condition is more contested. Note that it contains the weaker independence condition (see e.g., Dietrich 2007):

Independence. The result of the aggregation function for any proposition depends only on the judgements made on *this* proposition.

The intuitive plausibility of independence is easy to argue for (even though it is also not uncontested). Independence ensures that the collective judgement on a proposition is influenced only by individual positions on that specific proposition. If we consider a proposition P, changes in the profile regarding any other proposition should not influence the collective judgement on P.

Systematicity is more demanding than independence because it also requires that the same pattern of individual judgements on any proposition should lead to the same collective judgement on these propositions. More precisely, for any two propositions P, Q: if all individuals have the same judgements on P and on Q, then the collective results for P and Q must not differ. The intuitive ideal behind this condition is to treat all propositions equally. Systematicity rules out, for instance, the requirement of different qualified majorities for different propositions.

List and Pettit (2002) state and prove a theorem of judgement aggregation:

Theorem *There exists no judgement aggregation function generating complete, consistent and deductively closed collective sets of judgements which satisfies universal domain, anonymity and systematicity.*

This impossibility result has kicked off the research into questions of judgement aggregation and has led to a flourishing, often technically advanced literature (see List and Puppe 2009, for a survey). The theorem is important because it systematizes the special case of the discursive dilemma and shows that any form of judgement aggregation over a sufficiently complex agenda fails to meet all the described desiderata together. This poses a challenge for the aggregation of judgements: Either judgement aggregation fails (for some profiles), or one has to argue that at least one of the desiderata can and should be relaxed in order to avoid the impossibility result.[2]

Returning to the examples of the discursive dilemma above, I will now discuss four procedures to arrive at collective judgements: the majority vote on each proposition, the premise- and the conclusion-based procedure, and a dictatorship. The majority vote was already mentioned in the introduction of the discursive dilemma. If the collective votes on all propositions with simple majority, the group may end up with an inconsistent judgement set. This is unsatisfactory, and several ways to avoid this result have been proposed. The majority vote on all propositions satisfies universal domain, anonymity and systematicity, but fails to produce fully rational judgement sets for all logically possible judgement profiles.

[2]The literature on judgement aggregation has produced many refinements and extensions to List and Pettit's (2002) result, which cannot be described in detail here. Most important is perhaps Pauly and van Hees's (2006) generalizations, and further more general results in Dietrich and List (2007). The general structure of these additions is to discuss other, often weaker or differently constructed desiderata and prove impossibility (and sometimes possibility) results for aggregation functions. A very clear framework for judgement aggregation in general logic is provided by Dietrich (2007).

The premise-based procedure divides the propositions on the agenda into two sets: the premises and the conclusion(s). A majority vote is taken on each premise, and the premises adopted by these votes determine the remaining propositions, that is, the conclusions, by deductive closure. For the discursive dilemma stated in Table 3.1, P and $P \to Q$ can be taken as premises, Q as the conclusion. The majority adopts both premises, and deduces that Q must also be true. It therefore reaches the collective judgement set $\{P, P \to Q, Q\}$.[3] More loosely speaking, the premise-based procedure means: vote on the premises, deduce the conclusion. The premise-based procedure usually produces fully rational collective judgement sets,[4] but it violates systematicity because the collective judgement on the conclusion does not only depend on the individual judgements regarding the conclusion.

In the MI5 example, the worry with the premise-based procedure is that it overrules the majority vote. The second and perhaps more obvious procedure to the aggregation problem is to disregard the majority vote on the premises and only vote on the conclusion. This is the conclusion-based procedure. For Table 3.1 it leads the collective to adopt not-Q. Note that the collective does not take any view on the premises according to the conclusion-based view. Therefore, the conclusion-based procedure fails to produce complete collective judgement sets.

Another procedure to avoid collective inconsistency is to nominate a dictator, that is a person whose individual judgement set fully determines the collective judgement set. For instance, one could stipulate that the group always adopts the judgements of individual 1. Since the individual judgement sets are complete, consistent, and deductively closed, the "collective" judgement set will be, too. A dictatorship is a blatant violation of the anonymity condition, because a reshuffling of individuals (in particular, changing the dictator) may change the outcome.

Table 3.3 compares the four aggregation procedures. None of the procedures meets all the desiderata and the requirement of collective rationality (completeness, consistency, and deductive closure) together. List and Pettit's theorem shows that there is in fact no aggregation procedure that can meet all these desiderata together. It is therefore necessary to engage in a normative debate as to which desiderata should be sacrificed, or at least relaxed, to find a working aggregation procedure.

The desiderata under discussion are motivated by a broadly "democratic" set of values (see for instance List 2006). Universal domain can be normatively attractive from a democratic perspective because a democratically governed group should not rule out rational individual judgements ex ante. Anonymity can be attractive because it ensures that every member of the group has the same level of influence over the collective result. Systematicity ensures an equal treatment of all propositions, so that the aggregation procedure does not have an ex ante bias to define some

[3] It is not always the case that the propositions can be neatly divided into premises and conclusions. In addition, the premises do not necessarily determine the truth value(s) of the conclusion(s). For instance, if the votes on the premises had resulted in $\{\neg P, P \to Q\}$, the conclusion Q would not be determined by deductive closure because both Q and not-Q are consistent with the judgements on the premises.

[4] Except for those cases described in note 3.

Table 3.3 Aggregation procedures in comparison

Procedure	Universal domain	Anonymity	Systematicity	Collective rationality
Majority rule	+	+	+	–
Dictator	+	–	+	+
Premise-based	+	+	–	+
Conclusion-based	+	+	+*	–

*Systematicity holds for the conclusions

propositions as "special" or "more important". List and Pettit (2002) discuss several options to relax one of the three desiderata, or one of the three rationality conditions completeness, consistency, and deductive closure. Relaxing collective consistency and deductive closure is unattractive, because it results in irrational collective judgement sets. Other options are more attractive, depending on the circumstances. Relaxing universal domain is plausible when the individuals tend to have judgement profiles that are "well-behaved", that is do not give rise to the discursive dilemma. Relaxing anonymity may be justified when the competence in the group is unevenly distributed (List 2006). Relaxing systematicity is perhaps the most attractive move, because the idea that the collective judgements on different propositions do not influence each other appears implausible for a set of logically connected propositions in the first place. Even more implausible, systematicity also demands that all propositions are treated exactly equal in that regard. If a group deliberates on a number of dependent propositions, it should not be ruled out ex ante that the change of individual opinions on a proposition Φ can change the collective judgement on another proposition Ψ, even if the individual judgements on Ψ have not changed. Neither should it be ruled out that the same pattern of individual judgements for Φ and for Ψ can lead to different collective judgements on Φ and Ψ.

When considering distributed thinking systems, the background set of values to decide on an aggregation procedure does not necessarily have to be "democratic". But the properties one would like to see in a judgement aggregation function for distributed thinking may be similar. Universal domain is desirable from a distributed thinking perspective because the thinking process should not break down for certain inputs. Whether anonymity and systematicity are normatively desirable properties of a distributed thinking system is less clear. Anonymity is attractive if every thinking unit in the system of distributed thinking should be treated equally.[5] In the same vein, systematicity may be important when all propositions on the agenda should be treated equally.

Even if we relax one or more of the desiderata, we still need to say how we relax these desiderata and which aggregation functions we want to use. One important criterion for an aggregation function is that it meets the "knowledge challenge". This

[5] Also, relaxing anonymity does not yield particularly attractive aggregation procedures. In a very closely related setup, Pauly and van Hees (2006) show that the only aggregation procedure that meets all other desiderata is a dictatorship.

means that the aggregation function should be good at pooling the individual information that is distributed among individuals to reach correct outcomes. To explore the knowledge challenge in greater detail, I discuss the truth tracking performance of different aggregation functions.

An Epistemic Perspective

When voting on the truth or falsity of a single proposition, the Condorcet Jury Theorem shows that large groups can be almost always correct, as long as each member of the group is just slightly better than random at identifying the correct choice. Assume there is one correct state of the world, which is either that Φ or not-Φ is correct (or the better alternative). The competence assumption of the Condorcet Jury Theorem states that all individuals have a competence greater than 0.5. The competence of an individual is the probability to choose the correct alternative. With a competence greater than 0.5 the individuals are better than random in making the correct judgement between two alternatives.

The Condorcet Jury Theorem tells us: If all individuals have the same level of competence greater than 0.5, if their votes are independent,[6] and if they do not misrepresent their personal judgements for strategic reasons, then large groups will almost certainly choose the correct alternative in a majority vote.[7] The pooling of the individual competence in the vote renders the group much more competent than each single individual.

Let there be n individuals (with n being odd to avoid ties), and let the competence of all the different individuals 1 to n be p, with $p > 0.5$. The probability of a group to choose the correct alternative is (Grofman et al. 1983):

$$P^{CJT}(n, p) = \sum_{h=(n+1)/2}^{n} \binom{n}{h} p^h (1-p)^{n-h}. \qquad (3.1)$$

Table 3.4 shows the group competence for some levels of individual competence and different group sizes. One can see that even for relatively small levels of competence like 0.55, large groups reach a group competence of almost 100 %. Therefore, if the conditions of the Condorcet Jury Theorem hold, groups can be excellent "truth trackers" in dichotomous choice situations.

Figure 3.1 shows how the group competence develops for different group sizes and different values of p. One can see how larger groups quickly approach high competence if $p > 0.5$, but approach a group competence of 0 for $p < 0.5$.

[6] More precisely, if the votes are probabilistically independent, conditional on the truth value of Φ.

[7] The joint assumption of competence and independence rarely holds in practice. Weaker versions of the theorem have been proved (e.g. Dietrich and Spiekermann 2013). Dietrich (2008) points out that it is not possible to (statistically) justify both the independence and the competence assumptions and discusses less demanding assumptions and their implications.

Table 3.4 Group competence according to formula (3.1)

$n\|p$	0.501	0.51	0.55	0.6	0.7	0.8	0.9
11	0.503	0.527	0.633	0.753	0.922	0.988	≈1
101	0.508	0.580	0.844	0.979	≈1	≈1	≈1
1001	0.525	0.737	0.999	≈1	≈1	≈1	≈1

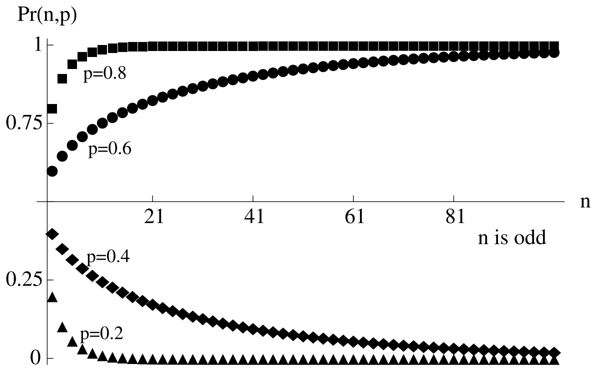


Fig. 3.1 Group competence $P^{CJT}(n, p)$ as a function of group size n, for different levels of individual competence p

The Condorcet Jury Theorem is the starting point to analyse richer collective decision problems. For the problems of judgement aggregation discussed above, each single proposition is a dichotomous choice problem, but the judgement aggregation problem as a whole is more complex. We have seen that there are different aggregation procedures, each with advantages and drawbacks. One possible normative criterion to decide for one aggregation procedure is to consider its epistemic performance, that is its ability to "track the truth". Here I focus primarily on a comparison between the conclusion- and the premise-based procedure, in line with discussions in Bovens and Rabinowicz (2006) and List (2006).

If a group follows the conclusion-based procedure, it simply votes on the conclusion, and disregards the premises. If the group follows the premise-based procedure, it votes on the premises and derives the conclusion by deductive closure. This will lead to different epistemic performances. I will show the diverging epistemic performances by discussing the detective example as stated in Table 3.2 above. The three proposition P, Q, R and their respective negations are on the agenda. In addition, all individuals accept $(P \wedge Q) \leftrightarrow R$ as true,[8] and we assume it is true as a matter of fact. Therefore, the world can be in 4 different states:

S1	P	Q	R
S2	$\neg P$	Q	$\neg R$
S3	P	$\neg Q$	$\neg R$
S4	$\neg P$	$\neg Q$	$\neg R$

[8] Assuming that the normative proposition R refers to a fact.

Less technically, all propositions can be true, or one of the premises and the conclusion are false, or both premises and the conclusion are false. These are also the only logically possible complete, consistent and deductively closed judgement sets, since we accept $(P \wedge Q) \leftrightarrow R$ as a background assumption.

For now, let us assume that a decision is epistemically correct if and only if it produces the correct stance on conclusion R (we discuss the idea that the stances on the premises should also be true below). In a premise-based procedure one votes only on the premises and deduces the conclusion. Therefore, the correct conclusion can be reached with different collective judgements on the premises:

State	Conclusion	Premise judgements with correct conclusion
S1	R	$\{P, Q\}$
S2	$\neg R$	$\{P, \neg Q\}, \{\neg P, Q\}, \{\neg P, \neg Q\}$
S3	$\neg R$	$\{P, \neg Q\}, \{\neg P, Q\}, \{\neg P, \neg Q\}$
S4	$\neg R$	$\{P, \neg Q\}, \{\neg P, Q\}, \{\neg P, \neg Q\}$

The point to note here is that the premise-based procedure can lead to the right conclusion even if one or both collective judgements on the premises are wrong. One can therefore be right for the wrong reasons. For instance, an agent can have the judgements P and $\neg Q$ and therefore $\neg R$, even though the world is in a state where $\neg P$ and Q are true. The agent is right to hold $\neg R$, but for the wrong reasons. If one wants the group to be right for "the right reasons" (Bovens and Rabinowicz 2006, p. 138f.), one should only consider cases where the collective judgements on both premises are correct, not only the conclusion derived from them.

I now turn to the conclusion-based procedure. There are two distinct ways for individuals to deal with a conclusion-based system. Either each single individual takes their judgements on the premises and derives the conclusion. This is the way Bovens and Rabinowicz propose. The conclusion-based procedure leads to a correct judgement on the conclusion if and only if a majority of individuals has the correct assessment of the conclusion. However, they may well have come to that assessment for the wrong reasons. For instance, if the correct assessment of the conclusion is that R is false, one can arrive at that conclusion from three different judgement sets on the two premises: $\{P, \neg Q\}$, $\{\neg P, Q\}$, $\{\neg P, \neg Q\}$. Only one set of judgements can be the right one, but all lead to the correct judgement on the conclusion. Alternatively, the agents completely disregard their judgements on the premises and make judgements only on the conclusion. In this case, the decision problem is collapsed into a decision on a single proposition, and the standard Condorcet Jury Theorem formula (3.1) applies. This way is unattractive from an epistemic standpoint because it completely disregards the information the individuals have on the premises.

Bovens and Rabinowicz calculate the probabilities for the group to make the right judgement on the conclusion. They consider four cases: (i) The use of the premise-based procedure where all correct conclusions are counted; (ii) the use of

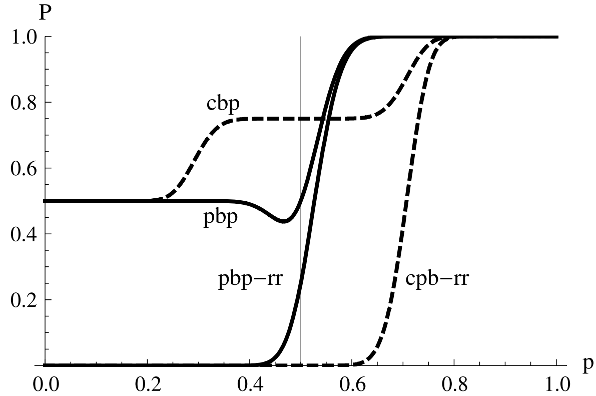


Fig. 3.2 Results for the premise- (*pbp*) and conclusion-based procedure (*cbp*) with $n = 101$, $\pi(P) = \pi(Q) = 0.5$. *rr* signifies the results is for the "rights reasons" constraint. The x-axis shows individual, the y-axis collective competence

the premise-based procedure where only judgements based on the right reasons are counted as correct; (iii) the use of the conclusion-based procedure where all correct conclusions are counted; and (iv) the use of the conclusion-based procedure where conclusions are counted as correct if a majority of individuals has reached the correct judgement on the conclusion for the right reason. These calculations depend on parameters. In addition to the individual competence p and the group size n, it also matters how likely the different states S1 to S4 are, which is determined by the prior probabilities of P, Q, and R. Let $\pi(P)$ be the prior probability that P is true; $\pi(Q)$ be the prior probability that Q is true. This in turn determines $\pi(R) = \pi(P)\pi(Q)$.

Figure 3.2 shows the results for the group competence, dependent on the individual competence p, for $n = 101$, $\pi(P) = \pi(Q) = 0.5$ and $\pi(R) = 0.25$. The two solid curves are the results for the premise-based procedure, the two dashed curves for the conclusion-based procedure. pbp is the result for the premise-based procedure, pbp-rr for the premise-based procedure when only results with the right reasons are counted as correct. Similarly, cbp shows the result for the conclusion-based procedure, cbp-rr the conclusion-based procedure with the correctness for the rights reason criterion.

First, consider the results for $p \geq 0.5$, that is the results with the (usually) more plausible assumption that individuals tend to be at least as good as a coin toss in making their decisions. For the premise-based procedure, the group competence is 0.5 for $p = 0.5$, and then quickly approaches 1 for larger p. If being right for the right reason matters, the group competence starts from a lower level, but still approaches 1 quickly. The conclusion-based procedure starts from a higher level (0.75), but is quickly outperformed by the premise-based procedure at a competence level of around 0.55. Interestingly, which procedure performs better depends on the level of p. Unsurprisingly, the procedures that care for being right for the right reasons have lower levels of collective competence. The premise-based procedure for the right reasons performs better than the conclusion-based procedure for the right reasons.

Now consider the results for competence levels lower than 0.5. First, there is a range of p for which the conclusion-based procedure fares better than the premise-based procedure. Second, the reliability of the premise-based procedure dips for

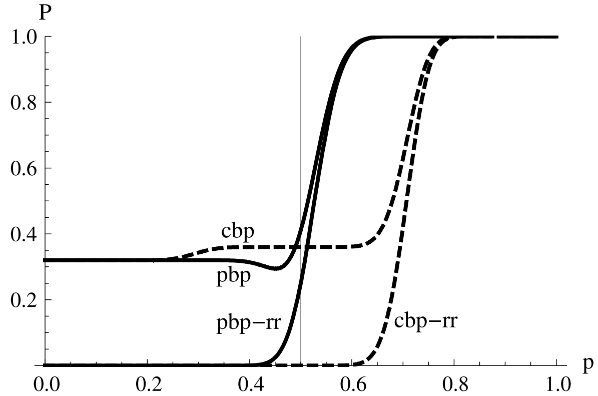


Fig. 3.3 Results for the premise- (pbp) and conclusion-based procedure (cbp) with $n = 101$, $\pi(P) = \pi(Q) = 0.8$. rr signifies the results is for the "rights reasons" constraint. The x-axis shows individual, the y-axis collective competence

values that are close but below 0.5 for p.[9] Third, if individuals are incompetent, they are very unlikely to be right for the right reasons. Overall, results for competence levels of $p < 0.5$ are of less interest because it is implausible that individuals are systematically worse than a coin toss.

Figure 3.3 shows the results for the same group size, but with different prior probabilities, namely $\pi(P) = \pi(Q) = 0.8$ and consequently $\pi(R) = 0.64$. For these parameter values, the premise-based procedure does better for all values $p > 0.5$. One can see that the performance of the two procedures depends on the prior probabilities. Both procedures perform worse around $p = 0.5$ compared to Fig. 3.2, but the conclusion-based procedure is still stronger in an area below 0.5.

List (2006, n. 25) criticizes the approach taken by Bovens and Rabinowicz because they do not distinguish between positive and negative reliability.[10] Positive reliability is the probability that the group correctly identifies R as true, negative reliability the probability that the group correctly identifies R as false. Different decision problems require different attention to the two reliabilities. Bovens and Rabinowicz simply calculate the probability that the group is correct. This may be misleading. Intuitively, this can be seen in Fig. 3.2 by considering the performances of the different aggregation procedures with $p = 0.5$, that is, when the individual competence is no better than a coin toss. An unbiased procedure should then be able

[9]This feature of the premise-based procedure has been overlooked by Bovens and Rabinowicz (see Fig. 6, where this dip is missing). The reason for this dip is quite easy to grasp intuitively: For very low p, the premise-based procedure is reliably wrong on both premises. If the world is in state S1 or S4, it will produce the wrong judgement on R, but if the world is in S2 or S3, it will produce the right outcome (though for the wrong reason, swapping the true and the false premise). As p approaches the watershed of 0.5, the procedure is less reliable false. It is still very unlikely that it is correct about both premises, but it is occasionally correct on one of them. Being sometimes right on one conclusion produces better results if the world is in S4, but worse results if the world is in either S2 or S3 (and it does not matter for S1). Since the world is more often in either S2 or S3 than in S4, the performance of the premise-based procedure dips for p close to but lower than 0.5.

[10]List also operates with asymmetrical individual competence, that is individuals have different competence for correctly judging true and false propositions.

to pick out the right result in half of the cases. But the conclusion-based procedure is doing much better. This is because the conclusion-based procedure has a bias towards assuming that R is false. Since Fig. 3.2 is drawn for $\pi(R) = 0.25$, this bias plays to the advantage of the conclusion-based formula. This result is due to a high negative reliability and the fact that for the given prior probabilities the conclusion is more often false than true. However, this comes at the cost of a low positive reliability. Figure 3.3 shows that the premise-based procedure is also biased for other parameter setting. Here both procedures show a bias that is to their disadvantage.

The upshot of the epistemic analysis is that different procedures for aggregating judgements have different qualities to "track the truth". These considerations show that a formal analysis of Goldman's "knowledge challenge" can help to decide which aggregation rule to use. For the example analysed here, the premise-based procedure performs well in most situations where individuals are competent. With regard to distributed thinking more generally, it is worthwhile exploring with formal models how different systems of distributed thinking lead to different epistemic success.

Distributed and Consistent Thinking

Distributed thinking can proceed in different ways. One way to conceptualize a distributed thinking system is to imagine a system where distributed non-thinking parts are connected in such a way that the whole assembly is a thinking system. A computer may be a distributed thinking system in that weak sense. Each single transistor could be seen as a non-thinking part, while the computer arranges these non-thinking parts in such a way that it can think, where thinking is taken as being able to solve logical problems. This notion of a thinking system is too weak because *any* thinking system is distributed in that sense. Brains, for instance, could be seen a distributed thinking system made of neurons.

The definition becomes more interesting if we assume that a distributed thinking system consists of several *thinking* sub-units. This definition is better because it rules out single computers and (perhaps) single brains, but includes relevant cases like groups of several agents, networked systems, etc. The interesting aspect of a distributed thinking system defined like that is the potential tension between the individual and the collective thinking. Oftentimes this tension is productive. We talk (rather vaguely) of "swarm intelligence" or "collective intelligence", and we sometimes experience how group deliberation can lead to better, more informed results than decisions by single individuals. But this tension can also lead to breakdowns of "collective intelligence", when no agreement can be reached, when the outcomes are inconsistent, or just plain wrong.

I have argued that judgement aggregation provides a useful framework for the analysis of distributed thinking. However, two anonymous referees argued that the judgement aggregation framework does not connect with the concept of distributed thinking for at least three reasons. First, judgement aggregation is *not dy-*

namic, in contrast to cognitive distribution, which emphasizes the dynamic interaction between the thinking units. Second, judgement aggregation does not engage with a central feature of the distributed cognition framework: the fact that *minds and world interact*, and that organisms reshape their environment in order to solve cognitive problems. This claim is often referred to by claiming that cognition happens "in the wild". Third, the judgement aggregation framework allegedly attempts to reduce distributed thinking to the thinking of sub-units, and does not appreciate that distributed thinking arises because *higher level structures emerge*.

My response is as follows. I largely concur with the first claim regarding the discursive dilemma, but point out that research in judgement aggregation raises interesting question about the possible dynamics that avoid the described impossibility results. This answer is connected with the second claim. While judgement aggregation per se does not address the interaction between minds and environment, it does raise questions as to how agents restructure their decision environment in order to avoid paradoxes like the discursive dilemma. Finally, I maintain that the validity of the third claim depends on the notion of emergence employed. In a weak sense, judgement aggregation and social choice theory support the claim that distributed thinking systems have emergent properties. I will now address each objection in greater detail.

Judgement aggregation, at least in the simple versions discussed here, does not incorporate a dynamic change of judgements through an interaction of individual and group judgements.[11] But the question of dynamics is raised indirectly by the impossibility results mentioned above, since the impossibility results pose the question how the breakdown of the aggregation process is avoided in practice. The discursive dilemma, for example, only arises for some of the many possible judgement profiles. It is therefore conceivable that a dynamic process, especially a process of deliberation, reduces or eradicates those profiles that lead to impossibility results. It is well known that the Arrow paradox can be avoided if the preference profile has certain structural properties, thereby relaxing the universal domain axiom (Dryzek and List 2003; Black 1948). Similar results hold for judgement aggregation. In case of the discursive dilemma, a suitable restriction of the universal domain axiom avoids the impossibility result (List and Pettit 2002). Empirical observations support the claim that deliberation leads to fewer occurrences of the discursive dilemma (List et al. 2013).[12] Thus, a dynamic process like deliberation may mitigate the occurrence of the impossibility result, and the framework of judgement aggregation raises interesting questions about the nature of the dynamic processes to avoid a breakdown of collective rationality. I therefore claim that even a static analysis in terms of judgement aggregation provides the debate on distributed thinking with useful concepts to analyse the dynamic processes.

[11] For judgement aggregation with regard to judgement *change* see List (2011).

[12] In addition, Bonnefon (2007) reports that individuals change their preference for the conclusion- and premise-based procedure with the nature of the decision.

The charge that judgement aggregation fails to scrutinize cognition "in the wild" can also be addressed by considering the escape routes to avoid impossibility results. Hutchins (1995) discusses several ways of how groups can structure their own decision making to simplify it, among them hierarchy and consensus (pp. 256–259). Clark also emphasizes the importance of "broader social and institutional contexts of action" (p. 186). List and Pettit show that if the individuals agree on a unidimensional alignment of the problem (similar to a left-right dimension in party politics), the dilemma can be avoided, even though individuals can disagree on their judgements. In addition, the dilemma disappears when the decision is delegated to specialists for each proposition (a form of "local dictatorship"), or when deliberation leads to a convergence of judgements. Thus, even though judgement aggregation does not directly explore group thinking "in the wild", the discussion of escape routes is very much concerned with the dynamic interaction of individuals and their potential to restructure the decision problem.

Finally, I turn to the charge that judgement aggregation is reductionist and fails to do justice to the emergent properties of distributed thinking systems. This charge hinges on the notion of emergence and reduction used. It is true that judgement aggregation is interested in aggregating individual to collective judgements. But the central result of the judgement aggregation research programme is that the aggregation is non-trivial, and that the group judgements cannot just be derived by summing up and counting the individual judgements. To underline this point, I use William Wimsatt's work on emergence and reduction. Wimsatt (1997) proposes a weak working definition for emergence: a system has emergent properties if it fails to be aggregative. For Wimsatt, the ideal aggregative system is invariant with regard to changes of like-for-like components, it scales linearly in size, the system properties are invariant with regard to a decomposition or reaggregation of the system, and there is no positive or negative interaction among the parts of the system. For instance, a heap of sugar is aggregative with regard to its mass. I can exchange one gram of sugar for another gram and its mass remains the same. If I add 1 gram of sugar, the total mass increases by 1 gram. If I divide the heap of sugar in two piles, the two piles each have half the mass of the original heap. If I put the heaps together again, I obtain the same mass. Finally, if I had two different types of sugar (brown and white sugar, say), this would not lead to positive or negative interactions in terms of the mass of the two types. Most systems are not entirely aggregative. For Wimsatt, the less aggregative a system is with regard to its properties, the more emergent properties it has.

Since the results presented above show that judgements on logically connected propositions cannot always be aggregated, given the stated axioms, such a system has emergent properties in Wimsatt's weak sense. One central result of the judgement aggregation research programme is that the sentence "A collective judgement of a group on a set of logically connected propositions is nothing but the aggregation of individual judgements" is not trivially true, since the aggregation encounters impossibility results. The results from judgement aggregation thus casts doubt on a simple "nothing but" reduction of group judgements, and weak emergence in Wimsatt's sense is embraced. For Wimsatt, "[a]n emergent property is—roughly—a system property which is dependent on the mode of the organization of the system's parts" (1997, p. S373, italics omitted). In this sense, the process of judgement aggregation has at least weak emergent properties. Whether this weak notion of emergence is enough to be of interest for the distributed cognition framework is a further question I leave to others. But I agree with Poirier and Chicoisne (2008) that the borders of distributed cognition are fuzzy.

Judgement aggregation as a field (in the simple treatments as discussed above) shows that even very simple reasoning processes run into difficulties when trying to turn rational individual judgements on logically connected propositions to rational collective judgements. If these problems arise even for the fairly simple problems like the discursive dilemma, one can anticipate similar and more difficult problems once one moves to more complicated settings. The basic lesson from the discursive dilemma is that the decision on the best processes applied in distributed thinking involves trade-offs between different properties of the reasoning process. Some processes are clearly worse than others, but when it comes to the best processes, different considerations need to be weighed against each other.

One possible consideration is the epistemic success of the procedures, i.e. the ability of the distributed thinking system to "track the truth". It is interesting to note that, for instance, in the comparison of the premise- and the conclusion-based procedure, it depends on the context of the decision problem which procedure performs best. However, if we introduce the additional requirement that the procedure must reach the correct decision for the right reasons, then the premise-based procedure is the clear winner in the example discussed. Being right for the right reasons can also be important if the group has to justify its decisions, or if the reasoning the group applies will be adopted or imitated in future reasoning processes.

Many extensions of the simple examples discussed in this chapter are possible. One should explore more complex decision problems, different logical dependencies, cases with incomplete judgement sets, heterogeneous competence levels, or settings where certain types of judgement errors are worse than others. Most of these settings have already been addressed in the literature on judgement aggregation and information pooling. The emerging literature on distributed thinking can benefit from the analytical and normative debates in these areas.

Acknowledgements An earlier version of this chapter was presented at the "Distributed Thinking" workshop at the University of Hertfordshire, organized by the Distributed Language Group. I am grateful for many constructive comments. I owe special thanks to Hemdat Lerman for several helpful discussions and extensive comments. The chapter benefited from the reviews provided by three anonymous referees and helpful comments from the editors of AI & Society. All remaining errors are my own.

Appendix

Bovens and Rabinowicz (2006) calculate the probabilities of the group being correct, conditional on the state. They define M^{pbp} as the proposition 'The premise-based procedure yields the correct result' and calculate probabilities conditional on all 4 states:

$$\begin{aligned}
P\big(M^{pbp}|\mathrm{S}1\big) &= P^{CJT}(n,p)^2,\\
P\big(M^{pbp}|\mathrm{S}2\big) &= Pr^{CJT}(n,p)^2 + P^{CJT}(n,p)\big(1-P^{CJT}(n,p)\big)\\
&\quad + \big(1-P^{CJT}(n,p)\big)^2,\\
P\big(M^{pbp}|\mathrm{S}3\big) &= P\big(M^{pbp}|\mathrm{S}2\big),\\
P\big(M^{pbp}|\mathrm{S}4\big) &= P^{CJT}(n,p)^2 + 2P^{CJT}(n,p)\big(1-P^{CJT}(n,p)\big).
\end{aligned} \tag{3.2}$$

Note that one can arrive at the correct result even though some or even both collective judgements on the premises are wrong. Given the logical dependency between the propositions, we know that $\pi(R) = \pi(P)\pi(Q)$. Summing up the conditional probabilities of being correct with the premise-based procedure, weighted by the probabilities that the different states obtain yields:

$$\begin{aligned}
P\big(M^{pbp}\big) = {} & P\big(M^{pbp}|\mathrm{S}1\big)\pi(P)\pi(Q) + P\big(M^{pbp}|\mathrm{S}2\big)\big(1-\pi(P)\big)\pi(Q)\\
& + P\big(M^{pbp}|\mathrm{S}3\big)\pi(P)\big(1-\pi(Q)\big)\\
& + P\big(M^{pbp}|\mathrm{S}4\big)\big(1-\pi(P)\big)\big(1-\pi(Q)\big).
\end{aligned} \tag{3.3}$$

Following Bovens and Rabinowicz's exposition for the conclusion-based procedure, let V be the proposition that a single voter determines the conclusion correctly, and $P(V)$ the probability the voter does so. Since each single voter applies deductive closure, we obtain the following probabilities for each single voter to be correct on the conclusion, based on their competence p:

$$\begin{aligned}
P(V|\mathrm{S}1) &= p^2\\
P(V|\mathrm{S}2) &= P(V|\mathrm{S}3) = p^2 + p(1-p) + (1-p)^2\\
P(V|\mathrm{S}4) &= p^2 + 2p(1-p).
\end{aligned} \tag{3.4}$$

Each individual can reach the correct conclusion by being correct on both premises (probability p^2) but one can also be correct, even if one is wrong on one or even both of the premises. Let M^{cbp} denote the proposition that the conclusion-based procedure yields the correct result. Conditional on the state, we can apply Eq. (3.1) to calculate the probability of a correct majority vote on the conclusion:

$$P\big(M^{cbp}|\mathrm{S}i\big) = P^{CJT}\big(n, P(V|\mathrm{S}i)\big). \tag{3.5}$$

Summing up the probabilities weighted by the prior probabilities of the different states yields:

$$\begin{aligned}
P\big(M^{cbp}\big) = {} & P\big(M^{cbp}|\mathrm{S}1\big)\pi(P)\pi(Q) + P\big(M^{cbp}|\mathrm{S}2\big)\big(1-\pi(P)\big)\pi(Q)\\
& + P\big(M^{cbp}|\mathrm{S}3\big)\pi(P)\big(1-\pi(Q)\big)\\
& + P\big(M^{cbp}|\mathrm{S}4\big)\big(1-\pi(P)\big)\big(1-\pi(Q)\big).
\end{aligned} \tag{3.6}$$

The results for the premise-based procedure in (3.3) and the conclusion-based procedure in (3.6) are based on the assumption that it does not matter whether the correct result is deduced from correct or incorrect judgements on the premises. If we want to be right "for the right reasons", the cases where incorrect judgements lead to correct outcomes need to be removed. Let M^{pbp-rr} denote the proposition that the group has arrived at the right judgement for the right reasons. This yields:

$$P\left(M^{pbp-rr}\right) = P^{CJT}(n, p)^2. \tag{3.7}$$

Similarly, for the conclusion-based procedure one want to consider the probability that a majority of voters is correct for the right reasons:

$$P\left(M^{cbp-rr}\right) = P^{CJT}\left(n, p^2\right).$$

References

Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). New York: Wiley.

Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy*, *56*, 23–34.

Bonnefon, J.-F. (2007). How do individuals solve the doctrinal paradox in collective decisions. *Psychological Science*, *18*, 753–755.

Bovens, L., & Rabinowicz, W. (2006). Democratic answers to complex questions: an epistemic perspective. *Synthese*, *150*, 131–153.

Clark, A. (1997). *Being there: putting brain, body, and world together again*. Cambridge: MIT Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7–19.

Dietrich, F. (2007). A generalised model of judgment aggregation. *Social Choice and Welfare*, *28*, 529–565.

Dietrich, F. (2008). The premises of Condorcet's jury theorem are not simultaneously justified. *Episteme*, *58*, 56–73.

Dietrich, F., & List, C. (2007). Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, *29*, 19–33.

Dietrich, F., & Spiekermann, K. (2013). Epistemic democracy with defensible premises. *Economics and Philosophy*, *29*(01), 87–120.

Dryzek, J., & List, C. (2003). Social choice theory and deliberative democracy: a reconciliation. *British Journal of Political Science*, *33*, 1–28.

Goldman, A. I. (2004). Group knowledge versus group rationality: two approaches to social epistemology. *Episteme*, *1*, 11–22.

Grofman, B., Owen, G., & Feld, S. L. (1983). Thirteen theorems in search of the truth. *Theory and Decision*, *15*, 261–278.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.

Kornhauser, L. A., & Sager, L. G. (1986). Unpacking the court. *The Yale Law Journal*, *96*, 82–117.

List, C. (2005). Group knowledge and group rationality: a judgment aggregation perspective. *Episteme*, *2*, 25–38.

List, C. (2006). The discursive dilemma and public reason. *Ethics*, *116*, 362–402.

List, C. (2011). Group communication and the transformation of judgments: an impossibility result. *Journal of Political Philosophy*, *19*(1), 1–27.

List, C. (2012). Collective wisdom: lessons from the theory of judgment aggregation. In H. Landemore & J. Elster (Eds.), *Collective wisdom: principles and mechanisms* (pp. 203–229).

List, C., & Pettit, P. (2002). Aggregating sets of judgments: an impossibility result. *Economics and Philosophy*, *18*, 89–110.

List, C., & Puppe, C. (2009). Judgment aggregation: a survey. In P. Anand, C. Puppe, & P. Pattanaik (Eds.), *Oxford handbook of rational and social choice*. Oxford: Oxford University Press.

List, C., Luskin, R. C., Fishkin, J. S., & McLean, I. (2013). Deliberation, single-peakedness, and the possibility of meaningful democracy: evidence from deliberative polls. *The Journal of Politics*, *75*(01), 80–95.

Pauly, M., & van Hees, M. (2006). Logical constraints on judgement aggregation. *Journal of Philosophical Logic*, *35*, 569–585.

Poirier, P., & Chicoisne, G. (2008). A framework for thinking about distributed cognition. In I. E. Dror & S. R. Harnad (Eds.), *Cognition distributed*. Amsterdam: Benjamins.

Vygotsky, L. S. (1962). *Thought and language*. Cambridge: MIT Press.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge: Harvard University Press.

Wimsatt, W. C. (1997). Aggregativity: reductive heuristics for finding emergence. *Philosophy of Science*, *64*, S372–S384.

Chapter 4
Computer-Mediated Trust in Self-interested Expert Recommendations

Jonathan Ben-Naim, Jean-François Bonnefon, Andreas Herzig, Sylvie Leblois, and Emiliano Lorini

Abstract Important decisions are often based on a distributed process of information processing, from a knowledge base that is itself distributed among agents. The simplest such situation is that where a decision-maker seeks the recommendations of experts. Because experts may have vested interests in the consequences of their recommendations, decision-makers usually seek the advice of experts they trust. Trust, however, is a commodity that is usually built through repeated face time and social interaction, and thus cannot easily be built in a global world where we have immediate Internet access to a vast pool of experts. In this article, we integrate findings from experimental psychology and formal tools from Artificial Intelligence to offer a preliminary roadmap for solving the problem of trust in this computer-mediated environment. We conclude the article by considering a diverse array of extended applications of such a solution.

Introduction

Important decisions are rarely made in isolation. Even when a single agent has the final say about what is to be done, the knowledge and information processing relevant to a complex decision are often distributed among several agents. Typically, one agent (the decision-maker) relies on one or several other agents (the experts) to provide recommendations based on their knowledge and know-how about the problem at hand.

The problem with experts, though, is that they may well have vested interests in the consequences of their recommendations. Think about investing. All of us who

J. Ben-Naim · J.-F. Bonnefon (✉) · A. Herzig · S. Leblois · E. Lorini
CNRS and Université de Toulouse, Toulouse, France
e-mail: bonnefon@univ-tlse2.fr

J.-F. Bonnefon · S. Leblois
CLLE, Maison de la Recherche, 5 al. Machado, 31058 Toulouse, France

J. Ben-Naim · A. Herzig · E. Lorini
IRIT, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse, France

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_4,

are not investment savvy might want to get some expert recommendations about what to do with our savings. Sometimes, our banker is willing to provide such recommendations. We are likely to take these recommendations with a grain of salt, though, because we are aware that the banker may have vested interests in pushing some specific financial products. We are facing a dilemma between our need for expert recommendation and the potentially self-interested character of the recommendations we can get from experts, who have vested interests in the decision we are going to make from their recommendation.

Our need for expertise thus makes us the potential targets of deception from self-interested experts. The traditional solution to this dilemma is to seek the recommendations of these experts and only these experts whom we endow with our *trust.* Trust is a multidimensional concept that has been informally defined as, e.g., "the expectation that the person is both competent and reliable, and will keep your best interest in mind" (Barber 1983), or, quite similarly, as a combined judgment of the integrity, ability, and benevolence of an individual (Mayer et al. 1995).

Trust is a commodity that is often built through repeated social interactions (Ferrin et al. 2006; Kramer 1999). Not only do people trust other people as a function of their interpersonal history, but even the most subtle aspects of face-to-face interaction can contribute to judgments of trustworthiness. For example, people are more ready to trust interaction partners who mimic their behavioral mannerisms (Maddux et al. 2008). Whether or not this is a sensible way to endow someone with trust is, of course, a debatable question. The point is, though, that behavioral mimicry requires face-to-face interaction, and that, generally speaking, feelings of trust commonly require a history of social interaction with the person whom is to be trusted.

This solution to the problem of trust is adapted to a small world where experts on a given topic are few and personally known to the decision maker. However, in our global village, an inexhaustible pool of experts on just any given topic is always just one mouse click away from us. Whatever our concern is, the Internet gives us a fast and convenient access to a vast number of experts. That would be good news, if only we knew which experts we could trust. The traditional solution to the problem of trust (repeated face time and social interaction) is no longer available in our global cognitive world.

In this chapter, we consider the problem of seeking expert advice through a web-based platform, where users are declared experts in various domains. We offer a list of suggestions for solving the problem of trust in this environment. The power of our approach resides in its multidisciplinarity, as we combine the cognitive insights of psychology to the formal methods of artificial intelligence to reach an integrated perspective on our problem. In the solution that we envision, regular users alternatively play the role of advisor or advisee in their interactions, depending on whom is in possession of the expert knowledge required by the situation. After each interaction, advisees have the possibility to appraise their advisor on the various dimensions that form the multifaceted concept of trust. The platform keeps a memory of these appraisals from which it can extract an aggregated, global index of the trustworthiness of any user, or decompose this global index into sub-indices corresponding to the various components of trust. Any new or regular user can thus attain a computer-mediated judgment of the extent to which any expert on the platform is to be trusted, or seek an expert whose detailed characteristics are optimally balanced to serve their needs.

In the rest of this chapter, we give a more detailed characterization of our problem, and we address in turn the various ingredients we need to sketch a solution. Section *Problem Specification* defines the problem of attaining a computer-mediated, complex judgment of trust within a multi-user, web-based platform of potentially self-interested experts. Section *Psychological Treatment* reviews experimental findings and psychological insights into the components of trust, their socio-cognitive antecedents, and their behavioral consequences. Section *Formal Treatment* builds on these materials and on the current state of the art in artificial intelligence to sketch a formal solution to our problem, which integrates the psychological constraints previously identified. Finally, section *Extended Applications* considers various extended applications of our suggestion for computer-mediated trust.

Problem Specification

Let us imagine that you came in possession of a banjo, which you would like to sell, but whose monetary value you have no idea of. One option would be to go to the closest (and probably the only) banjo store you know of, and to ask the owner to appraise your banjo. The problem, though, is that the owner is not only the person whom you can ask about the value of your banjo, but also the person you are likely to sell the banjo to. Not knowing whether you can trust the owner not to take advantage of the situation, you turn to a web-based platform for musical instruments amateurs, where you are likely to find plenty of users who can appraise a banjo, and plenty of potential banjo buyers. Your trust problem, though, is just demultiplicated, because these are likely to be broadly the same persons. The fact that you now have an abundance of experts you might solicit is no improvement over your previous situation, because you do not have the time, the resources, or the motivation to engage in repeated social interactions with all these people in order to find out who you can trust.

We believe that the platform should offer a solution to achieve the same results as repeated social interaction. It should provide you with the basic parameters that form the building blocks of trust, as well as some index of the extent to which you can trust your potential advisors. We believe this service can be achieved by formalizing the notion of trust, and taking advantage of the history of the advisor-advisee interactions on the platform.

Not every user of the platform is an expert of everything. To continue our musical instruments example, some users may declare expertise in appraising banjos, whilst others may declare expertise in appraising cellos. Thus, depending on the situation, a given user may be in a position to give expert advice, or to receive it. Now consider that everytime a user x receives expert advice from another user y, x is given the opportunity to appraise this advice on all the dimensions that the complex notion of trust is known to encompass. The platform records this interaction as a tuple $R_{xy}\langle r_1, \ldots, r_n \rangle$, where $r_1, \ldots, r_n$ are the appraisals given by x about the recommendation of y on the various dimensions of trust. Soon enough, the platform should be in a position to answer a request about the trustworthiness of agent y, by aggregating the information contained in the tuples expressed about y.

A number of problems must be solved to achieve such a result. First, we need to decide on the exact nature of the appraisals $r_1, \ldots, r_n$. Then, we need to decide on the way these ratings should be aggregated, both at the individual level and at the collective level. Finally, we need a formal characterization of all the components of trust and of the properties one can used to reason about them, in order to generalize our solution to environments where artificial agents interact with human agents, or among themselves. Solving these problems requires a multidisciplinary approach, drawing on experimental psychology as well as artificial intelligence methods. We now consider in turn the insights given by these two disciplines.

Psychological Treatment

Various definitions of trust can be found in the psychological literature. Some authors define trust mostly in terms of its behavioral consequences, e.g., 'Trust is the extent to which a person is confident in, and willing to act on the basis of, the words, actions, and decisions of another' (McAllister 1995), or trust is 'the willingness to accept vulnerability based upon positive expectations about another's behavior' (Rousseau et al. 1998). Early structural perspective on trust distinguished between trust based on cognition and trust based on affect (Johnson-George and Swap 1982; Lewis and Weigert 1985; Rempel et al. 1985). 'Cognitive' trust is based on explicit knowledge and 'good reasons' to think that a person is reliable or dependable. 'Affective' trust is based on an emotional bond between individuals. Clearly, just as behavioral mimicry, emotional bonds are not within the scope of our application. We should thus focus on that sort of trust which is based on explicit knowledge and deliberative thought.

Idealilly suited for our purpose is the suggestion that trustworthiness is a three-dimensional attribute composed of competence, benevolence, and integrity (Barber 1983; Mayer et al. 1995). Competence reflects the ability of a person with respect to the task at hand. Benevolence reflects a positive attitude towards the truster, and a genuine concern for the truster's interests. Integrity reflects the adherence of the trustee to an appropriate set of ethical principles. Let us now consider in turn these three components of trust, and their potential importance in situation where advice is given.

Competence

Many studies have investigated the influence of an advisor's perceived competence on the uptake of her recommendations. Perhaps unsurprisingly, these studies concur that the recommendation of an advisor is more influential when her perceived expertise is greater. Interestingly, people seem ready to accept claims of expertise at face value, even in experimental situations where the quality of the offered ‘expert’ advice is actually weak (Harvey and Fischer 1997). While it is clear why people seek the advice of individuals they believe to be more competent than they are, we note that people are sometimes ready to seek the advice of individuals they believe to be *less* competent than they are; in particular, when the stakes of the decision are serious enough that they want to share the responsibility for the decision, whatever the relative expertise of their advisor (Harvey and Fischer 1997).

People appear to use a variety of cues to appraise the expertise or competence of an advisor. For example, advisors who express high confidence in their recommendation are perceived as more competent, and their recommendation is given more weight by the decision maker (Sniezek and Buckley 1995; Van Swol and Sniezek 2001). Likewise, advisors who give very precise recommendations (as opposed to vague estimates) are perceived as more competent, and, again, their recommendation is given more weight by the decision maker (Yaniv and Kleinberger 2000). All other things being equal, these strategies do appear to increase the quality of the decision making, for there seems to be an ecologically valid correlation between expertise, confidence, and precision (Sniezek and Buckley 1995; Van Swol and Sniezek 2001; Yaniv et al. 1991). Then again, these studies did not control for the possibility that the advisor has vested interests in the decision of the advisee; and a self-interested advisor may well express a very precise recommendation with great confidence, only to better serve her own interests.

Finally, a reputation for expertise is hard to build, but rapidly destroyed (Slovic 1993; Yaniv and Kleinberger 2000). Many useful recommendations are required before one is trusted as a competent advisor, but only a few average or bad recommendations are enough to lose that reputation. This phenomenon can be related to the more general *negativity bias* in impression formation (Ito and Cacioppo 2005; Skowronski and Carlston 1989). The negativity bias refers to the greater weight we attribute to negative behaviors when inferring personality traits: For example, fewer negative behaviors are needed to infer a negative trait, compared with the number of positive behaviors we need to infer a positive trait. The negativity bias, and its specific consequences for the dynamics of trust, can be conceived as a safeguard for a species that exhibits a strong tendency to spontaneous cooperation, ensuring that untrustworthy partners are quickly detected and unprofitable cooperation promptly forsaken.

Benevolence

Whenever a conflict of interest is possible, and even when it is not, people are concerned about the degree to which their advisors really care about their interests. A benevolent advisor genuinely cares about the best interests of the advisee, has a positive attitude towards the advisee, and thinks about the advisee’s interests at least as much as her owns.

Even when the advisor has no explicit vested interest in the situation, benevolence can contribute to trustworthiness independently of competence. For example, the mere fact that the advisee already knows the advisor (a proxy for benevolence) makes a difference to the advisor's perceived trustworthiness, even when controlling for the advisor's expressed confidence in her advice (a proxy for competence); in fact, this expressed confidence no longer affects trustworthiness as soon as the advisor and the advisee know each other (Van Swol and Sniezek 2001). In these experiments, an increase in trustworthiness translated into a greater weight put on the advisor's recommendation. Other experimental studies directly made it clear to decision makers whether or not some advisor was benevolent, concerned about their best interests. These experiments concurred that recommendations from benevolent advisors are given greater weight in the decision (White 2005).

Interestingly, it has been claimed that people are ready to trade off competence for benevolence when the emotional load of their decision is high (White 2005). One experiment put subjects in a situation to decide whether they would leave their savings in a badly performing fund, or take them out. In the low emotional load condition, the savings were meant to pay for a summer band camp for young musicians. In that case, subjects sought competent rather than benevolent advisors. In the high emotional load condition, the savings were meant to pay for college. In that case, subjects sought benevolent advisors, and were ready to sacrifice some level of competence in order to ensure benevolence.

Whether this effect is truly due to emotional load or to another confounded variable is not quite clear, but the possibility of a trade off between competence and benevolence would already be especially relevant to our current purpose, given that we conceptualise competence and benevolence as different dimensions of the complex concept of trust. It would mean that a global index of trust might not be precise enough to accommodate people's needs. Indeed, different situations may require different mix of competence and benevolence, although the global index of trust would remain the same.

Benevolence-based trust can obviously be harmed by malevolent behavior. However, it can be repaired on the long term by subsequent benevolent behavior, or, on the short term, by promises to adopt a benevolent behavior. Apologies for malevolent behavior do not seem sufficient, though, to repair trust (Schweitzer et al. 2006).

Integrity

The integrity of the advisor reflects her unconditional adherence to a set of principles deemed appropriate by the advisee. Note that integrity so defined can be independent of benevolence. For example, one may expect an advisor to maintain confidentiality whether or not one believes the advisor to be benevolent. Conversely, one may question whether an advisor can be trusted to maintain confidentiality, independently of whether this advisor is benevolent or not.

Some indices of trust put a strong emphasis of integrity. For example, recent studies investigating the relation between emotion, trust, and the uptake of advice (Dunn and Schweitzer 2005; Gino and Schweitzer 2008) used a measure of trust that focused on whether the advisor could be expected to unconditionally honor commitments, and whether the advisor could be expected to unconditionally tell the truth. These studies found that incidental emotions (i.e., which were felt independently of the advisor) could affect this integrity-based trust, which affected in turn the weight given to the advisor's recommendation. More specifically, incidental anger decreased integrity-based trust, and incidental gratitude or happiness increased integrity-based trust.

Finally, integrity-based trust seems hard to repair once it has been harmed by a failure to honor one's commitment (Schweitzer et al. 2006). Once an individual has failed to deliver on a promise, her trustworthiness appears to be durably impaired, and not significantly repaired by apologies or renewed promises to change behavior, even when these promises are genuinely honored.

Summary

Agents faced with difficult decisions often find that they do not possess all the knowledge and expertise required to make the best possible choice. A natural solution is then to seek expert recommendation about the decision; but because experts may have vested interests in the consequences of their recommendation, they need to be trusted by the agent making the decision. Our global world offers easy access to a vast pool of experts; but it does not offer the traditional guarantees of trustworthiness that come with a history of personal interaction with all these experts.

This problem of computer-mediated trust in expert recommendations clearly falls within the scope of the distributed cognition framework proposed by Hollan and collaborators (Hollan et al. 2000). Indeed, it presents the three following characteristic features:

- Cognitive processes are distributed across the members of the social group. Not only is the final decision codependent on computations made by the decision maker and by the expert, but the trust granted to the expert is itself the result of distributed computation among the users of the platform.
- Cognitive processes involve coordination between internal and external structure. To reach an overall assessment of trustworthiness, the decision maker cannot simply inquire into the judgments made by others, but must delegate some computations to the platform and coordinate with the results of this computation.
- Processes are distributed through time in such a way that the products of earlier events transform the nature of later events. Indeed, the dynamics of trust is such that events cannot be interpreted in isolation. A display of integrity, for example, has a very different impact on trustworthiness depending on whether the expert is known to have given at least one dishonest recommendation.

Overall, the computer-mediated construction of trust is a distributed cognitive process exhibiting a complex trajectory over agents, events, and time, and requires coordination with an external computational structure. It does not result, however, in any radical conceptual rewiring of the nature of mind or trust. In that sense, we offer a 'weak' distributed perspective, focused on the multi-level aggregation of the cognitive outputs of humans and artefacts: a formally difficult problem, but a tractable one.

Our approach sticks to a conceptualisation of the mind as an information processing system, with clearly defined inputs and outputs; and our work rests on the assumption that a significant portion (though clearly not the whole) of trust-building boils down to information processing. Although some aspects of trust-building elude our formalization, we believe that the cold information processes captured by our formalization can already offer some solid decisional grounds. These processes are constrained by a number of variables and psychological dimensions, which we explored in the previous section. In line with previous psychological research, we conceptualise trust as a multidimensional concept comprising competence (expert ability), benevolence (positive attitude and concern towards the interests of the advisee), and integrity (unconditional adherence to a set of principles deemed appropriate by the advisee).

These three dimensions of trust exhibit different degrees of asymmetry in the differential impact of positive and negative information. In the case of integrity, negative information receives extremely greater weight than positive information. This asymmetry is also observed with respect to competence, but apparently to a lesser extent. Finally, the asymmetry would appear to be the least pronounced in the case of benevolence.

Some compensation seems possible between the dimensions of competence and benevolence, since situations appear to exist where advisees are willing to sacrifice some measure of competence to ensure some measure of benevolence. It is less clear whether integrity can be traded that way, or whether it should be considered as a completely non-compensatory dimension of trust. One possibility, that would need empirical validation, is that the level of integrity functions as the upper-bound for the level of trustworthiness. A related solution to the problem of computer-mediated trust is to first filter out advisors who have been judged to lack integrity; and then to provide the user with aggregated indices of competence and benevolence, without taking the responsibility to trade one for another in a global index of trustworthiness. This responsibility should be left to the user, who knows best whether the situation primarily calls for competence, benevolence, or both.

We now turn to the formal treatment of our problem. We introduce a logical framework wherein the three aspects of trust can be formally characterized, and wherein we can model trust reasoning about these three aspects.

Formal Treatment

This section presents a logical framework called $\mathcal{TRUST}$ in which the competence, benevolence and integrity of an advisor can be formally characterized. $\mathcal{TRUST}$ is a multi-modal logic which supports reasoning about time, agents' actions and agents' mental attitudes including beliefs and goals. It also allows to express the normative concept of obligation. In this sense, $\mathcal{TRUST}$ combines the expressiveness of dynamic logic (Harel et al. 2000), temporal logic (Emerson 1990) and deontic logic (Åqvist 2002) with the expressiveness of a so-called BDI (belief, desire, intention) logic of agents' mental attitudes (Cohen and Levesque 1990). We introduced the logic $\mathcal{TRUST}$ in our previous works on the logical formalization of the concepts of trust and reputation (Lorini and Demolombe 2008). It is not the aim of this work to discuss the precise semantics of the modal operators of the logic $\mathcal{TRUST}$. We just present them in an informal way by highlighting their intuitive meanings and their basic properties.[1]

The syntactic primitives of the logic $\mathcal{TRUST}$ are the following:

- a nonempty finite set of agents $AGT = \{i, j, \ldots\}$;
- a nonempty finite set of atomic actions $AT = \{a, b, \ldots\}$;
- a finite set of propositional atoms $ATM = \{p, q, \ldots\}$.

The language of $\mathcal{TRUST}$ is defined as the smallest superset of ATM such that:

- if $\varphi, \psi \in \mathcal{L}$, $\alpha \in ACT$ and $i \in AGT$ then $\neg\varphi$, $\varphi \vee \psi$, $\texttt{Does}_{i:\alpha}\varphi$, $\texttt{Bel}_i\varphi$, $\texttt{Choice}_i\varphi$, $\texttt{Past}\varphi$, $\texttt{Obl}\varphi \in \mathcal{L}$.

ACT is the set of complex actions and is defined as follows:

$$ACT = AT \cup \left\{ inf_j(\varphi) | j \in AGT, \varphi \in \mathcal{L} \right\}.$$

An action of the form $inf_j(\varphi)$ denotes the action of informing agent j that φ is true. We call this kind of actions informative actions.

Thus, the logic $\mathcal{TRUST}$ has five types of modalities: $\texttt{Bel}_i$, $\texttt{Choice}_i$, $\texttt{Does}_{i:\alpha}$, $\texttt{Past}\varphi$ and $\texttt{Obl}$. These modalities have the following intuitive meaning.

- $\texttt{Bel}_i\varphi$: the agent i believes that φ;
- $\texttt{Does}_{i:\alpha}\varphi$: agent i is going to do α and φ will be true afterward ($\texttt{Does}_{i:\alpha}\top$ is read: agent i is going to do α);
- $\texttt{Past}\varphi$: it has at some time been the case that φ;
- $\texttt{Choice}_i\varphi$: agent i has the chosen goal that φ holds (or simply agent i wants that φ holds).

Operators of the form $\texttt{Choice}_i$ are used to denote an agent's chosen goals, that is, the goals that the agent has decided to pursue. We do not consider how an agent's chosen goals originate through deliberation from more primitive motivational attitudes called desires (see e.g. Rao and Georgeff 1991; Conte and Castelfranchi 1995; Castelfranchi and Paglieri 2007 on this issue).

The following abbreviations will be convenient:

[1] See for instance Lorini and Demolombe (2008) for an analysis of the semantics of these operators, their relationships, and their correspondence with the structural conditions on the models of the logic $\mathcal{TRUST}$.

$$\texttt{Intends}_i(\alpha) \stackrel{\text{def}}{=} \texttt{Choice}_i \texttt{Does}_{i:\alpha} \top$$

$$\texttt{Inf}_{i,j}(\varphi) \stackrel{\text{def}}{=} \texttt{Does}_{i:inf_j(\varphi)} \top$$

$$\texttt{BelIf}_i \varphi \stackrel{\text{def}}{=} \texttt{Bel}_i \varphi \vee \texttt{Bel}_i \neg\varphi.$$

$\texttt{Intends}_i(\alpha)$ stands for 'agent i intends to do action α'. This means that i's intention to perform action α is defined by agent i's choice to perform action α. $\texttt{Inf}_{i,j}(\varphi)$ stands for 'agent i informs agent j that the fact φ is true'. Finally, $\texttt{BelIf}_i \varphi$ stands for 'agent i believes whether φ is true'.

Operators for actions of type $\texttt{Does}_{i:\alpha}$ are normal modal operators satisfying the axioms and rules of inference of the basic normal modal logic K (Chellas 1980).

Operators of type $\texttt{Bel}_i \varphi$ are just standard doxastic operators in Hintikka style (Hintikka 1962) satisfying the axioms and rules of inference of the so-called system KD45 (Chellas 1980). It follows that an agent cannot have inconsistent beliefs, and an agent has positive and negative introspection over his beliefs. Formally:

$\mathbf{D}_{Bel}$ $\neg(\texttt{Bel}_i \varphi \wedge \texttt{Bel}_i \neg\varphi)$
$\mathbf{4}_{Bel}$ $\texttt{Bel}_i \varphi \rightarrow \texttt{Bel}_i \texttt{Bel}_i \varphi$
$\mathbf{5}_{Bel}$ $\neg\texttt{Bel}_i \varphi \rightarrow \texttt{Bel}_i \neg\texttt{Bel}_i \varphi$

As emphasized above, operators of the form $\texttt{Choice}_i$ express an agent's chosen goals. These are similar to the modal operators studied in Cohen and Levesque (1990). Since an agent's chosen goals result from the agent's deliberation, they must satisfy two fundamental rationality principles: chosen goals have to be consistent (i.e., a rational agent cannot decide to pursue inconsistent state of affairs); chosen goals have to be compatible with the agent's beliefs (i.e., a rational agent cannot decide to pursue something that it believes to be impossible). Thus, every operator $\texttt{Choice}_i$ is supposed to satisfy the axioms and rules of inference of the so-called system KD Chellas (1980). It follows that an agent cannot choose φ and $\neg\varphi$ at the same time. Moreover chosen goals have to be compatible with beliefs. Formally:

$\mathbf{D}_{Choice}$ $\neg(\texttt{Choice}_i \varphi \wedge \texttt{Choice}_i \neg\varphi)$
$\mathbf{Comp}_{Bel,Choice}$ $\texttt{Bel}_i \varphi \rightarrow \neg\texttt{Choice}_i \neg\varphi$

As far as the modal operator for obligation is concerned, we take the operator of Standard Deontic Logic (SDL) (Åqvist 2002). That is, the modality $\texttt{Obl}$ is also supposed to satisfy the axioms and rules of inference of the so-called system KD. It follows that obligations have to be consistent. That is:

$\mathbf{D}_{Obl}$ $\neg(\texttt{Obl}\varphi \wedge \texttt{Obl}\neg\varphi)$

The temporal operator $\texttt{Past}\varphi$ is also a normal modality which satisfies the axioms and rules of inference of system of the basic normal modal logic K. The following two additional axioms are added in order to capture two essential aspects of time.

$\mathbf{4}_{Past}$ $\texttt{Past}\texttt{Past}\varphi \rightarrow \texttt{Past}\varphi$
$\mathbf{Connected}_{Past}$ $(\texttt{Past}\varphi \wedge \texttt{Past}\psi) \rightarrow (\texttt{Past}(\varphi \wedge \texttt{Past}\psi) \vee \texttt{Past}(\psi \wedge \texttt{Past}\varphi) \vee \texttt{Past}(\psi \wedge \varphi))$

Axiom $\mathbf{4}_{Past}$ says that the past satisfies transitivity: if it has been the case that it has been the case that φ then it has been the case that φ. Axiom $\mathbf{Connected}_{Past}$ just expresses that the past is connected: if there are two past moments t and t' then either t is in the past of t' or t' is in the past t or $t = t'$.

Other relationships between the different modalities of the logic $\mathcal{TRUST}$ are expressed by the following logical axioms.

$\mathbf{Alt}_{Does}$ $\quad \texttt{Does}_{i:\alpha}\varphi \rightarrow \neg\texttt{Does}_{j:\beta}\neg\varphi$
IntAct $\quad \texttt{Does}_{i:\alpha}\top \rightarrow \texttt{Intends}_i(\alpha)$
$\mathbf{Inc}_{Time,Does}$ $\quad (\varphi \wedge \texttt{Does}_{i:\alpha}\top) \rightarrow \texttt{Does}_{i:\alpha}\texttt{Past}\varphi$

Axiom $\mathbf{Alt}_{Does}$ says that: if agent i is going to do α and φ will be true afterward, then it cannot be the case that agent j is going to do β and $\neg\varphi$ will be true afterward. Axiom **IntAct** relates an agent's intentions with his actions. According to this axiom, an agent is going to do action α only if has the intention to perform action α. In this sense it is supposed that an agent's *doing* is by definition intentional. A similar axiom has been studied in Lorini and Herzig (2008), Lorini et al. (2006) in which a logical model of the relationships between intention and action performance is proposed. Finally Axiom $\mathbf{Inc}_{Time,Does}$ expresses that every action occurrence goes from the present to the future (i.e. actions do not go back to the past). That is, if φ is true in the present and agent i does action α then, after the occurrence of action α, φ is true at some point in the past.

Formal Definitions of Competence, Benevolence and Integrity

The aim of this section is to formalize in the logic $\mathcal{TRUST}$ the three properties competence, benevolence and integrity of a potential advisor.

We start with the concept of competence of an advisor to provide good recommendations about a certain issue φ.

Definition 1 (Competence) Agent j is a competent advisor (or competent information source) about a certain issue φ if and only if, if agent j believes that φ then φ is true.

This notion of competence can be formally expressed as follows:

$$\texttt{Competent}_j(\varphi) \stackrel{\text{def}}{=} \texttt{Bel}_j\varphi \rightarrow \varphi.$$

The second concept we aim at formalizing is benevolence.

Definition 2 (Benevolence) Agent j is a benevolent advisor (or benevolent information source) about a certain issue φ if and only if, for every agent i, if j believes that i wants to believe whether φ is true and j believes that φ is true then j informs i about his opinion.

This notion of benevolence can be formally expressed as follows:

$$\texttt{Benevolent}_j(\varphi) \stackrel{\text{def}}{=} \bigwedge_{i \in AGT} \big((\texttt{Bel}_j\texttt{Choice}_i\texttt{BelIf}_i\varphi \wedge \texttt{Bel}_j\varphi) \rightarrow \texttt{Inf}_{j,i}(\varphi)\big).$$

As far as integrity is concerned, we split this concept into three different concepts of sincerity, confidentiality and obedience. That is, we suppose that the expression 'the advisor satisfies the property of integrity' means that the advisor is sincere, obedient, and he guarantees the confidentiality of the information.

Definition 3 (Sincerity) Agent j is a sincere advisor (or sincere information source) about a certain issue φ if and only if, for every agent i, if j informs i that φ is true then j believes that φ is true.

This notion of sincerity can be formally expressed as follows:

$$\texttt{Sincere}_j(\varphi) \stackrel{\text{def}}{=} \bigwedge_{i \in AGT} \big(\texttt{Inf}_{j,i}(\varphi) \rightarrow \texttt{Bel}_j\varphi\big).$$

Definition 4 (Confidentiality (or Privacy)) Agent j is an advisor (or information source) which guarantees the confidentiality (or privacy) of the information φ if and only if, for every agent i, if it is obligatory that j does not inform i that φ is true then j does not inform i that φ is true.

This notion of confidentiality can be formally expressed as follows:

$$\texttt{Privacy}_j(\varphi) \stackrel{\text{def}}{=} \bigwedge_{i \in AGT} \big(\texttt{Obl}\neg\texttt{Inf}_{j,i}(\varphi) \rightarrow \neg\texttt{Inf}_{j,i}(\varphi)\big).$$

Definition 5 (Obedience) Agent j is an obedient advisor (or obedient information source) about a certain issue φ if and only if, for every agent i, if j is obliged to inform i about φ then j informs i about φ.

This notion of obedience can be formally expressed as follows:

$$\texttt{Obedient}_j(\varphi) \stackrel{\text{def}}{=} \bigwedge_{i \in AGT} \big(\texttt{Obl Inf}_{j,i}(\varphi) \rightarrow \texttt{Inf}_{j,i}(\varphi)\big).$$

We define the integrity of the advisor j about a certain issue φ as the logical conjunction of j's sincerity about φ, j's obedience about φ, the fact that j guarantees the confidentiality of the information φ:

$$\texttt{Integrity}_j(\varphi) \stackrel{\text{def}}{=} \texttt{Sincere}_j(\varphi) \wedge \texttt{Privacy}_j(\varphi) \wedge \texttt{Obedient}_j(\varphi).$$

Trust Reasoning About Competence, Benevolence and Integrity

When assessing the trustworthiness of a certain advisor k, the truster i evaluates whether k has the three properties of competence, benevolence and integrity. In many situations, such an evaluation might depend on what agent i has heard about the advisor k in the past. In particular, agent i's evaluation of an agent k's competence, benevolence and integrity might be based on what the other agents told to i about k. In these situations, agent i has to apply certain procedures for *aggregating* all information that he has received from the other agents about k's properties.

The logic $\mathcal{TRUST}$ allows to formalize some of these procedures, namely *majority* and *unanimity*. For instance, we can specify the concept of 'the majority of agents informed agent i that agent k is benevolent about φ',

$$\begin{aligned}&\mathtt{Maj}_i\big(\mathtt{Benevolent}_k(\varphi)\big)\\&\quad\stackrel{\text{def}}{=}\bigvee_{J\subseteq AGT,|J|>|AGT\setminus J|}\Big(\bigwedge_{j\in J}\mathtt{Past}\ \mathtt{Inf}_{j,i}\big(\mathtt{Benevolent}_k(\varphi)\big)\Big).\end{aligned}$$

According to this definition, the majority of agents informed agent i that agent k is benevolent about φ (noted $\mathtt{Maj}_i(\mathtt{Benevolent}_k(\varphi))$) if and only if there exists a group of agents J such that every agent j in J informed i that k is benevolent about φ and J is larger than its complement with respect to AGT.

In a similar way we can express that 'the majority of agents informed agent i that agent k is competent about φ',

$$\begin{aligned}&\mathtt{Maj}_i\big(\mathtt{Competent}_k(\varphi)\big)\\&\quad\stackrel{\text{def}}{=}\bigvee_{J\subseteq AGT,|J|>|AGT\setminus J|}\Big(\bigwedge_{j\in J}\mathtt{Past}\ \mathtt{Inf}_{j,i}\big(\mathtt{Competent}_k(\varphi)\big)\Big).\end{aligned}$$

As far as unanimity is concerned, we can specify the concept of 'all agents unanimously informed agent i that agent k satisfies the property of integrity',

$$\mathtt{Unan}_i\big(\mathtt{Integrity}_k(\varphi)\big)\stackrel{\text{def}}{=}\bigwedge_{j\in AGT}\mathtt{Past}\ \mathtt{Inf}_{j,i}\big(\mathtt{Integrity}_k(\varphi)\big).$$

The previous definitions of majority-based benevolence and competence and unanimity-based integrity can be used to specify the procedures adopted by agent i to evaluate a certain advisor k. From the experimental literature that we reviewed in section *Psychological Treatment*, it seems sensible to use a strong unanimity procedure for integrity, but to allow a more lenient majority procedure for competence and benevolence:

$$\mathtt{Maj}_i\big(\mathtt{Competent}_k(\varphi)\big)\rightarrow\mathtt{Bel}_i\mathtt{Competent}_k(\varphi).$$

This rule says that if the majority of agents informed i that k is a competent advisor then i believes so,

$$\mathtt{Maj}_i\big(\mathtt{Benevolent}_k(\varphi)\big)\rightarrow\mathtt{Bel}_i\mathtt{Benevolent}_k(\varphi).$$

This rule says that if the majority of agents informed i that k is a benevolent advisor then i believes so,

$$\mathtt{Unan}_i\big(\mathtt{Integrity}_k(\varphi)\big)\rightarrow\mathtt{Bel}_i\mathtt{Integrity}_k(\varphi).$$

This rule just says that if all agents informed i that k is an advisor which satisfies the property of integrity then i believes so.

At this point, and although much has still to be articulated, we will conclude the formal analysis of our problem. Indeed, our goal in this chapter has not been to solve the problem of computer-mediated trust in partial expert recommendations, but rather to provide a roadmap for addressing this problem, by integrating findings from experimental psychology and formal tools from Artificial Intelligence. In the last section of this chapter, we go beyond our initial problem by suggesting extended applications of our approach, to a range of problems where trust (or reputation) cannot be assessed by personal interaction, where agents cannot be vouched for by an objective arbiter, but where the possibility remains of applying some variant of our approach.

Extended Applications

In its most general formulation, the problem we have tackled here concerns multi-agent applications where users have to evaluate (or simply compare) agents, but it is impossible to call on an objective arbiter to provide some help. This may happen for various reasons, for example, the number of agents is too large, no arbiter is considered sufficiently competent and sincere, arbiters are too expensive, etc. However, in such applications, a lot of feedbacks may be available, that is, information about agents provided by other agents. Trust and reputation systems of the kind we have envisioned here are conceived to exploit such information in order to help users to take decisions about other agents.

The information provided by peers should be used with caution. It can be incomplete, and it may be downright false. Indeed, agents may have vested interests in their judgments, and therefore may lie or hide the truth to serve their interests. Another issue that is critical in any trust and reputation system is that of *cycles* of information (e.g., a provides information about b, b provides information c, and c provides information a). Trust and reputation systems have to give different weights to the pieces of information provided by the agents, but assigning such weights in a rational way turns out to be difficult in the presence of information cycles. In this final section, we consider several situations where a trust and reputation system can be used to overcome the absence of a neutral, objective arbiter.

Currently, the best-known examples of a virtual community of agents are social networks such as of Facebook or MySpace. We briefly evoke this setting because of its popularity, although it does not, strictly speaking, relate to our topic; indeed, the reputation system that can be implemented in this setting is likely to be gratuitous (it is not meant as a decision help) and unrelated to our central issue of trust. Still, in such a social network, a wealth of information is given by agents about other agents. For example, in addition to the comments and pictures they leave on each other 'walls', users can rate their virtual friends on a number of dimensions (are they attractive? honest? serious?), or vote for the nicest person in their network; and all this information can be used to extract aggregated judgments about any particular user.

Other applications are, to a greater extent, geared to help decisions. For example, e-commerce applications like Ebay are such systems where it is useful to have information about sellers before deciding to buy an item. Here, the agents are the users and the dimension of trust that is the most decisive is integrity, the expectation that the seller respects his commitments and tell the truth. In this kind of system, there are too many buyers and sellers for an external arbiter to evaluate them all. However, after each transaction, buyer and seller have an opportunity to appraise each other. This rich amount of feedback can be exploited to reach aggregated evaluation of individual ebayers. Ebay is already equipped with a simple reputation system, which does not however explicitly measure a score of integrity-based trust. Rather, it uses a simple scheme where a positive feedback from a buyer brings one point, a negative feedback removes one point, and a neutral feedback has no consequences. Symbolic trinkets are attached to some scores (e.g., a star when the seller reaches a net score of 10 points). One limitation of this system is that it does not weight feedback according to the reputation of the ebayer who provided it.

Agents in a trust and reputation system need not be human. Indeed, web pages may be seen as agents, and a link between two pages may be construed as a positive recommendation by the linking page about the linked page. Pages that gathered the most aggregated support can be considered as more trustworthy along the competence dimension of trust: they are the pages were relevant information is to be found. This is in fact one of the broad principles that PageRank (the reputation system used by the Google search engine) is based on.

Scientific citation indices offer another application of trust and reputation systems, where scientific papers are the agents (or, perhaps, the minions of the scientists who wrote them). In most scientific reputation indices, citing a paper is construed as a positive recommendation of that paper. This is true of very basic indices such as the raw number of citations, as well as of more elaborated indices such as the h index. As often, this framework gives every citation the same weight in the aggregated evaluation of the paper or the scientist who wrote the collection of papers under consideration. A trust and reputation system would allow to weight a citation according to aggregated scores of the citing paper that would take into account the potential for vested interests in citing one article rather than another.

The last application we consider in this discussion is less publicized, partly because of its more technical nature. It concerns the important issue of message encryption and public key certificates. Without engaging in too technical a discussion, we can summarize the problem as follows. Various agents wish to exchange encrypted messages. Each agent is in possession of two personal keys. One of these is public, it can be used to encrypt messages sent to this agent; the other is private, it is used by the agent to decrypt messages that were encrypted using his or her public key.

One concern within this framework is that a malicious agent may assume the identity of another agent a, and pretend that her own public key is actually the public key of a. Other agents may then mistakenly believe they are encrypting messages with the public key of a, when they are really using the public key of the malicious agent. The malicious agent can then intercept and decrypt messages that were meant for a.

The problem for any agent, then, is to have sufficient ground to believe that what is advertized as the public key of a truly is the public key of a. Public key certificates are used to solve that problem. A public key certificate is a document supposedly written by an agent b, signed with a public key K_b, that certifies that an agent a is the owner of a public key K_a. Consequently, we can extract from this framework a set of pairs composed of an agent and a public key supposedly belonging to it. In addition, we can extract a binary support relation between these pairs. More precisely, we consider that a pair (b, K_b) supports the credibility of a pair (a, K_a) if there exists a certificate from b, signed with K_b, stating that a is the owner of K_a. This information can be used to evaluated the credibility of the different pairs. For example, the well-known Pretty Good Privacy system looks for chains of pairs, where the credibility of the first element is trusted by the sender, each element supports the next one, and the last element is the receiver.

The main limitation of this framework is its extreme cautiousness. If the chain of certification does not go back to some agent trusted by the potential sender, no encrypted message can be sent. One way to overcome this extreme cautiousness (at some risk), is to use the kind of trust and reputation system that we have considered through this article, and to appraise the trustworthiness of an agent-key pair based on the structure of the certification graph.

We do not consider this application in any greater detail, for our goal in this last section was rather to give a broad perspective of the various problems that can be tackled by the general approach we have outlined in this chapter. We hope that the reader will have gained a sense of the many domains where a trust and reputation system can help appraise the characteristics of some agents who cannot be evaluated by a central, neutral authority. These applications must be supported by a mix of psychological findings and artificial intelligence formalisms, whose exact composition depends on the extent to which the agents in the system are human or human-like in their behavior and intentions.

References

Åqvist, L. (2002). Deontic logic. In D. M. Gabbay & F. Geunther (Eds.), *Handbook of philosophical logic*. Dordrecht: Kluwer Academic.

Barber, B. (1983). *The logic and limits of trust*. New Brunswick: Rutgers University Press.

Castelfranchi, C., & Paglieri, F. (2007). The role of beliefs in goal dynamics: prolegomena to a constructive theory of intentions. *Synthese*, *155*, 237–263.

Chellas, B. F. (1980). *Modal logic: an introduction*. Cambridge: Cambridge University Press.

Cohen, P. R., & Levesque, H. J. (1990). Intention is choice with commitment. *Artificial Intelligence*, *42*, 213–261.

Conte, R., & Castelfranchi, C. (1995). *Cognitive and social action*. London: London University College of London Press.

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: the influence of emotion on trust. *Journal of Personality and Social Psychology*, *88*, 736–748.

Emerson, E. A. (1990). Temporal and modal logic. In J. van Leeuwen (Ed.), *Handbook of theoretical computer science, volume B: formal models and semantics*. Amsterdam/Cambridge: North-Holland/MIT Press.

Ferrin, D. L., Dirks, K. T., & Shah, P. P. (2006). Direct and indirect effects of third-party relationships on interpersonal trust. *Journal of Applied Psychology*, *91*, 870–883.

Gino, F., & Schweitzer, M. E. (2008). Blinded by anger or feeling the love: how emotions influence advice taking. *Journal of Applied Psychology*, *93*, 1165–1173.

Harel, D., Kozen, D., & Tiuryn, J. (2000). *Dynamic logic*. Cambridge: MIT Press.

Harvey, N., & Fischer, I. (1997). Taking advice: accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*, 117–133.

Hintikka, J. (1962). *Knowledge and belief*. New York: Cornell University Press.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, *7*, 174–196.

Ito, T. A., & Cacioppo, J. T. (2005). Variations on a human universal: individual differences in positivity offset and negativity bias. *Cognition and Emotion*, *19*, 1–26.

Johnson-George, C. E., & Swap, W. C. (1982). Measurement of specific interpersonal trust: construction and validation of a scale to assess trust in a specific other. *Journal of Personality and Social Psychology*, *43*, 1306–1317.

Kramer, R. M. (1999). Trust and distrust in organizations: emerging perspectives, enduring questions. *Annual Review of Psychology*, *50*, 569–598.

Lewis, J. D., & Weigert, A. (1985). Trust as social reality. *Social Forces 63*, 967–985.

Lorini, E., & Demolombe, R. (2008). Trust and norms in the context of computer security. In *LNCS: Vol. 5076. Proc. of the ninth international conference on deontic logic in computer science (DEON'08)* (pp. 50–64). Berlin: Springer.

Lorini, E., & Herzig, A. (2008) A logic of intention and attempt. *Synthese*, *163*(1), 45–77.

Lorini, E., Herzig, A., & Castelfranchi, C. (2006). Introducing "attempt" in a modal logic of intentional action. In *LNAI: Vol. 4160. Logics in artificial intelligence: 10th European conference (JELIA 2006)* (pp. 280–292). Berlin: Springer.

Maddux, W. W., Mullen, E., & Galinsky, A. D. (2008). Chameleons bake bigger pies and take bigger pieces: strategic behavioral mimicry facilitates negotiation outcomes. *Journal of Experimental Social Psychology*, *44*, 461–468.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*, 709–734.

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*, 24–59.

Rao, A. S., & Georgeff, M. P. (1991). Modelling rational agents within a BDI-architecture. In *Proceedings of the 2nd international conference on principles of knowledge representation and reasoning (KR'91)* (pp. 473–484). San Mateo: Morgan Kaufmann.

Rempel, J. K., Holmes, J. G., & Zanna, M. D. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, *49*, 95–112.

Rousseau, M., Sitkin, S., Burt, R., & Camerer, C. (1998). Not so different after all: a cross-discipline view of trust. *The Academy of Management Review*, *23*, 393–404.

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: restoring violated trust. *Organizational Behavior and Human Decision Processes*, *101*, 1–19.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: a review of explanations. *Psychological Bulletin*, *105*, 131–142.

Slovic, P. (1993). Perceived risk, trust and democracy. *Risk Analysis*, *13*, 675–685.

Sniezek, J. A., & Buckley, B. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes*, *62*, 159–174.

Van Swol, L. M., & Sniezek, J. A. (2001). Trust, confidence and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, *84*, 288–307.

White, T. B. (2005). Consumer trust and advice acceptance: the moderating roles of benevolence, expertise, and negative emotions. *Journal of Consumer Psychology*, *15*, 141–148.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*, 260–281.

Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin*, *110*, 611–617.

Chapter 5
Living as Languaging: Distributed Knowledge in Living Beings

Anton Markoš, Jana Švorcová, and Josef Lhotský

Abstract We trace life at different levels of organization and/or description: from protein ecosystems in the cell up to the cohabitation of individuals within and between historically established lineages. Ways of such cohabitation depend on experience of particular guilds or aggregates; they cannot be easily foretold from any basic level of description, they are distributed across all levels, and across all members of the community. Such phenomena of interactivity constitute a lived world which, we argue, represents a genuine analogy with domains of human cultures and languages. We draw an analogy with three levels of meaning as defined by Rappaport (Ritual and religion in the making of humanity, Cambridge University Press, Cambridge, 2010) and make an attempt to show that life and languaging are virtually analogous.

Contributions to this volume show that cognition arises not only 'in the head', but also as the result of living in a network of interactions—in the medium of languaging; language and languages cannot be separated from languaging (Steffensen 2013), and our joint activities make sense because of how we concert our doings in a culture or what Thibault (2011) terms a social meshwork. Outcomes of such doings often depend also on differences that people find as meaningful cues to perform expertly or to construe wordings in a particular way. In other words, much depends on patterns that are extracted by living beings that dwell in a historical world of bodily experience, and of the community into which they are rooted. Indeed, in the context, these ideas will not seem controversial; however, in what follows, we propose taking a further step: we propose that analogical processes help *all* inhabitants of the biosphere/semiosphere to become valuable members of such living networks. Our

A. Markoš (✉) · J. Švorcová · J. Lhotský
Faculty of Science, Department of Philosophy and History of Science, Charles University Prague, Viničná 7, Praha 2, 128 44, Czech Republic
e-mail: markos@natur.cuni.cz

J. Švorcová
e-mail: jana.svorcova@natur.cuni.cz

J. Lhotský
e-mail: lhotsky@natur.cuni.cz

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_5, © Springer-Verlag London 2013

approach may look as yet another contribution to the long list of holistic theories that compete without success with the reigning reductionist paradigm of biology. We, however, do not deny the explanatory power of contemporary biological theory: by stressing the role of historical bodily experience and of the "cultural" role of communities we strive towards a fuller understanding of life phenomena, much along the line the linguists undertook from the vocabularies through semiotics up to languaging. We invite the reader to take an excursion from the "central dogma" and neo-Darwinian explanation of evolution, towards what we believe is a more complete view of the living, that extends through 9 orders of magnitudes (or "73 octaves of nature's music," as poetically expressed by Ho 1993) and from nanoseconds to 4 billions of years. Our extension to the distributed view is to argue that what goes for cognition and language also applies generally to life.

Levels of Meaning

Biologists have much to gain from considering how human cultures exploit what have been termed various 'levels' of meaning. Here, we take inspiration and a leading thread in the book by Rappaport (2010), *Ritual and religion in the making of humanity*; we shall exploit its paraphrase "Ritual in the making of species," still by following Rappaport's argumentation that was intended for the human race only.

Rappaport invites us to acknowledge human cultures as featuring three levels of meaning. Our paper will take the view that the kinds of systems that we find in molecular biology bear remarkable similarities. (1) *Low-order meaning* is based in differences that can be found in the everyday semantics: thus *rat* differs from both *mouse* and *rate* (in spelling as in pronunciation). Plainly, science is most comfortable with this kind of meaning, and we shall investigate some features of this level in biology. (2) In the *middle-order* of meaning, a person is able to make and construe "similarities hidden beneath the surfaces of apparently distinctive phenomena" (Rappaport 2010, p. 71). While types may still appear, they are now associated with various kinds of metaphors and signs. This is the level of biosemiotics and biohermeneutics, and we took casual examples how an individual construes its body and its umwelt at this level of meaning. Finally, (3) *high-order meaning* is "grounded in identity or unity, the radical identification or unification of self with other" (p. 71); in dealing with this, we look beyond models that depend on the regular appearance of discrete types and draw on what we think of as "experience of being" and our sense of belonging in a community. Rappaport concludes (*caveat lector*, he speaks about human condition!): "The distinctions of low order meaning, lodged in language, divide the world into discrete objects; the recognition of similarity constituting middle-order meaning makes connections among those objects; high-order meaning unifies the world into wholeness. Middle-order and high-order meaning may thus prevail, at least from time to time, over the experiences of fragmentation and alienation that are, possibly, concomitants of language's objectifying powers, but it is important to note that *the three levels of meaning do not always live easily together*. Naive scientism and naive rationalism tend to deny the validity of middle- and high-order meaning, and it is ironically interesting that information may be the enemy of meaningfulness. Conversely, untempered commitment to middle- and high-order meaning may ignore crucial distinctions in the natural and social world." (Rappaport 2010, p. 73)

Let us explain the three levels on a Biblical parable: Ezequiel cites the Lord as declaring: "I have no pleasure in the death of the wicked" (33, 11). While the verse may be new to some readers, it has been cited and interpreted in numerous sermons, moral debates and literary contexts. Yet, we suggest, none of hypothetical readers, naïve or learned, is likely to have considered the sentence in terms of the following syllogism:

p: God has no pleasure in the death of the wicked
q: Mrs. A is wicked
$p \rightarrow q$: Mrs. A is immortal

Yet this is what the sentence means in plain English! If language functioned like a unidirectional code, it would evoke Rappaport's low-order level of meaning. Why, then, do we not attribute immortality to Mrs. A? Our case is that the networks or paraphernalia of our civilization leads us to read the verse in relation to higher orders of meaning. This is not based on interpretation of the discrete signs at all: we *feel* that God would not grant such things; our cognitive biases link "death" with "damnation." Readers who are familiar with the Babylonian exile will place the prophet's words yet in another context that derives from our understanding of the original, our acquaintance with Middle Eastern realia, and our own cultural milieu. Still nothing of this will explain, why 2600 years after being written, the verses do still appeal to people in, for example, Central Russia or Arizona.

Middle- and high-level layers of meaning are often seen as bound up with hermeneutics, therefore as part of the humanities, as distinguishing humans from the rest of the biosphere. Biology, it is assumed, can be studied independently of history, cultural contexts, language-like patterns, experience and so on. Yet, in "The chaos of everyday life," Rappaport suggests (2010, xvi), "stability is bound up with the social facts of a shared collective existence." Not only do we depend on history, the reiteration of forms and experience but we also draw on, clichés, metaphors, ritualized activities and even strange assumptions. In Umberto Eco's terms: "… it is impossible to communicate without putting something into the *background frame of mutual agreement and assuming that the other is able to access this presupposed knowledge*. Otherwise, each speech event would require a complete restatement, with the result that there would be no time to say, or listen to, anything. This is clearly too great an extension for a presupposition as a sentence phenomenon, since the utterance of even the simplest sentence can presuppose all the world in this sense." (Eco 1994, pp. 228–229, emphasis added.)

In turning to how language and cognition play out 'in the wild', such ways of meaning appear less exotic. Human actions are situated in a normative world where bodies use learning (and other interactional products) to co-ordinate internal and external structure. People, moreover, do this collectively as 'co-acting assemblages' (Cowley and Vallée-Tourangeau 2010, p. 469). Persons-embedded-in-action and/or-interaction resemble to a "field force," built and rebuilt continuously by inhabitants of a "field" that was inherited from those who long since passed away. Heidegger (1982 [1995]) calls this Being-with-others (*Mitsein*) in a Country (*Gegend*). This country is moulded by, on the one side, tectonic forces and, on the other, efforts by those who share their being in the here and now. In this way, a countryside or culture is able to evolve across innumerable generations. If this is indeed the basis of cognition, it cannot be traced to simple encodings. This is because, in coming up with thoughts, people draw on distant factors—like the words of an ancient prophet in our example above. Can we, however, generalize from human experience to biosphere, without committing the flaw of anthropomorphization?

In what follows, we show that biological codes such as the DNA script, intracellular and intercellular signal systems and ecological cohabitations also have a strange duality. While participating in unidirectional processes, they also inhabit a 'country' of messages and lineages. The "scientific" treatment of "biological syllogisms" applies in artificial, laboratory settings. Like thinking and sense-making, living processes and their evolution depend on interactivity, a process Kirsh (2010, p. 441) defines as a "back and forth process where a person alters the outside world, the changed world alters the person, and the dynamic continues." Many challenge such a view: in line with the central dogma of molecular biology (see below), many focus on the lowest Rappaportian level. It is hoped that higher levels of meaning are emergent phenomena that can be explained by focusing on such a lowest level of description. It is as if, in studying life, one can ignore the role of living beings. Yet, in Western culture and, thus, the humanities, this view is common; even Rappaport (2010, p. 10), who should know better, concurs: "Non-human systems are organic systems constituted largely by genetically encoded information. Human systems are cultural-organism systems constituted by symbolic (linguistic) as well as genetic information." In challenging this, we aim to rescue the study of life itself from the no man's land that lies between sciences and humanities. Our axiom be: *All living systems are cultural-organism systems constituted by symbolic* (*linguistic*) *as well as genetic information*.

The Low Order

Central dogma of molecular biology: Genetic information inscribed sequentially in nucleic acids (DNA or RNA) is decisive for the structure and function of proteins; proteins, however, have no means for feedback, that is, they cannot implement any changes in the genetic information. As proteins represent the principal agency in construction of the body (the phenotype), it follows, that all relevant information how to build a body is inscribed in the form of a linear string of building blocks ("characters") constituting the chain of nucleic acids.

In biology, by adopting the so called "Central dogma," the focus has fallen exclusively on the low order of meaning (see any modern textbook in molecular or cell biology, e.g., Alberts et al. 2008; the reader who is acquainted with basics of molecular biology can safely skip this section). It claims that information flow in biological systems is unidirectional, from script encoded in DNA to proteins to higher levels of organization. Hence, the basic level of description of any living being is its master copy of DNA containing "data" and "programs" how to build the body. Even instructions how to construct the "hardware" (or better, wetware) of the body must be in its entirety encoded in the genetic script (its "wording" is called *genotype*). In the process of *transcription*, parts of DNA information (about 30,000 "genes," constituting about 2–4 per cent of DNA in human cells) are transferred to much shorter strings of RNA; one class of RNAs (messenger or mRNAs) serves for *translation* of information into a string of a particular *protein* (more about proteins see below). The translation rules—the genetic code—extend the realm of chemistry: the code was established in evolution by *natural conventions* (Barbieri 2008a, 2008b). Several thousands of different kinds of proteins constitute the lowest level of cellular agency (higher levels being multiprotein complexes, membranes, organelles, and other structures) responsible for metabolism, locomotion, cell division, but above all for the extremely reliable *replication*, that is copying of DNA master copy, to distribute it to daughter cells, and, of course, also "to read" it in the transcription and translation processes described above. The assembly of agencies and structures constitute cells, and cells build multicellular bodies; how such an assembly looks like, that is what is its *phenotype*, is *primarily* the function of the genetic script implemented in DNA. To repeat again: there is *no reverse flow of information*—no feedback from the world or flesh into the script (see, e.g., Monod's 1972 classical treatise *Chance and necessity*). Phenotypes, and other structures of biosphere web, essentially obey, as if verbatim, the genotypic instructions. There are no pterodactyls in contemporary biosphere, because no pterodactyl genotypes operate in contemporary cells, they were lost long ago. Flaws and paradoxes of the theory came to light relatively early (see, e.g., Hofstadter 1979), yet the debates on the topic often end with a mantra "In principle the central dogma holds." The problem, of course, lies in the fact that all living beings have been born of living beings, they do not start from scratch like crystals, flames, neither are they products of assembly lines. Bodies and they genetic scripts are co-extensive, neither is "primary" or more basic.

The contemporary neo-Darwinian paradigm, however, draws on the Central dogma. Replication of DNA is highly, but not absolutely reliable (typos, and even more serious disruptions may occur due also to external factors), hence, genes in a population may come out in slightly different "spellings" called *alleles* (likewise, "program" and "programme" represent two alleles of the same word). Different alleles (and coalitions of alleles) result in proteins, that is also bodies (phenotypes), slightly differing (in this or that respect) from other individuals present in the population, and such differences may influence the *fitness* of that particular body—in terms of the amount of its descendants. The body is, then, a vehicle to transmit its burden of its alleles into the next generation: the fitter the vehicle, the higher the frequency of particular allele(s) in the population in the next generation. The fitness is determined by *natural selection* in the external environment: Because of the Central dogma, natural selection acts on the carnal vehicles, whereas the gist of evolution is in transferring pure information as inscribed in DNA. For a more succulent version of the story see e.g. Dawkins (1976, 1982).

What is important for our further exploration is the fact that the Central dogma and neo-Darwinism models presuppose the concept of a 'basic level' in description of the living. Living beings are viewed as passive machines that are designed to transfer their "software" into their progeny. One way of countering such views is to exploit the language metaphor of life (Markoš and Faltýnek 2011; Markoš et al. 2009; Kleisner and Markoš 2005, 2009). Rather than dwell at the lowest level of meaning, we look beyond models that depend on discrete types and, in so doing, show the relevance of higher levels of meaning to the realm of living. As we argue below, living systems draw also on ecological (or oiko-logical) aspects of meaning. It is our view that recognition of their historical basis is necessary to placing life in a coherent system of knowledge that brings out the continuities that link it with the many human worlds that unfold within a cultural meshwork. As there is no external agency steering the living processes and their evolution, we argue that life acts as its own designer (Markoš et al. 2009), that is, the lowest level of meaning will not satisfy the task. Yet in other words: the agency driving both ontogeny and evolution is distributed across many levels of bodily organization, with no primary, or central, steering controls.

The Middle Level

One way of countering central dogma with its basic level of description is to exploit the language metaphor of life (Markoš and Faltýnek 2011; Markoš et al. 2009; Kleisner and Markoš 2005, 2009). Rather than dwell at the lowest level of meaning, we look beyond models that depend on discrete types and, in so doing, show the relevance of higher levels of meaning to the realm of living. In so doing, we face opposition from both the sciences and the humanities (see, e.g., Heidegger 1982 [1995]). However, we see no need for this: accordingly the paper aims to show that, in contrast to views associated with the kind of logic associated with the central dogma, living systems draw on ecological (or oiko-logical) aspects of meaning. It is our view that recognition of their historical basis is necessary to placing life in a coherent system of knowledge that brings out the continuities that link it with the many human worlds that unfold within a cultural meshwork.

We pursue the "language metaphor of life" beyond the affairs of *Homo sapiens*, into communities of living entities. In arguing that it is essential to show that history and experience matter to intracellular processes, cells living in a body, members of a species and even ecosystems. Life depends on, or better dwells in, cultures or, in Kauffman's (2000) terms, *biospheres* made up of populations of cooperating *autonomous agents*. Many of the properties of languaging (Markoš 2002; Markoš et al. 2009; Markoš and Švorcová 2009; Markoš and Faltýnek 2011) appear in communities or guilds of living entities: the processes that sustain life are *radically distributed in that they depend on 'memory' that is inseparable from their surroundings*. Living beings are not produced or created *ex nihilo* like crystals or tornadoes: they are *born* into already present "biospheric fields." Parental individuals (and the community) give birth to beings that develop in a pre-existing domain of rules, values, heuristics and ways of doing things. Hence, besides the genetic script and the body that harbors its patterns, we emphasize the third factor—the community. We now focus our approach around four examples: (1) the intracellular "ecology" of the protein world; (2) epigenetics; (3) symbiosis and symbiogenesis; and (4) the new science of evolution-development, affectionately known as evo-devo (Carroll 2005; Gilbert and Epel 2009).

Proteins as Agents at the Molecular Level

In our view it is difficult to understand life without considering properties of the protein community. Proteins are huge molecules. By comparison, if we treat the "size" of a hydrogen atom as 1 and that of water as 18, a protein averages at about 40,000 (10–100,000). Each of their "building blocks," an amino acid, has a size of around 100. Proteins are always synthesized as linear chains consisting of aperiodic sequences that are constituted by 20 different species of amino acids; the chain is synthesized according to a sequence of particular sections in DNA called genes. Genes are first copied (transcription) into "messenger RNA" which is translated in accordance with a sequence of instructions (the genetic code) into the amino acid chain that constitutes the protein. It should be pointed out that the whole process is catalyzed and steered by pre-existing protein "machinery" that, in its turn, arose also from the transcription-translation process.

The resulting native protein chain shows sensitivity to a particular train of amino acids by wrapping onto itself and creating a 3D molecular protein molecule. Given the view that all necessary information is contained in the DNA (e.g., Monod 1972; Anfinsen 1973) many thought that a one-dimensional codon sequence unequivocally determined both the chain and the shape of the molecule. On this view, since proteins are the "basic building blocks" of the cell, the shape of cells and multicellular beings is to be traced to the code of a genetic script. In fact, a protein molecule can attain an astronomic number of different shapes. In a given case, however, their embedding in a cellular environment will ensure that only a limited ("meaningful") number are attained (Fig. 5.1). Misfoldings are quickly repaired, or removed—by the cell's protein assembly apparatus.

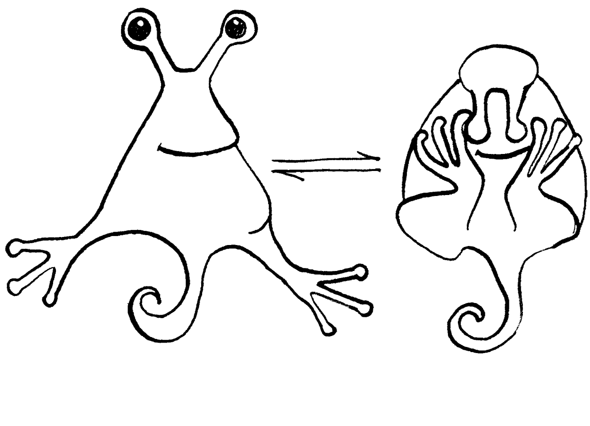


Fig. 5.1 Two possible conformations of a protein molecule

Most proteins possess *binding site*(*s*) for a specific *ligand* (a small molecule or specific shapes recognized on macromolecules like DNA, RNA, sugars, or other proteins). On binding the ligand, the molecule *does* something by *changing its shape* (conformation): it may change the chemical nature of the ligand (enzyme), bind an antigen (antibody), transfer molecules across a membrane (channels, pumps); pass or amplify signals (receptor); etc. These are not coding processes (based on input-output relations) but rather *performances* that change the protein molecule's shape while binding its ligand(s). Every protein depends on being able to change its shape upon interacting with its environment. A mammalian cell contains about 30,000 genes of which, in a given cell, 10,000 are typically 'read'. However, the set of actual protein shapes in the cell is much larger: as explained below, this depends on the protein ecosystem into which new proteins are born (for more detail and self-explanatory cartoons, see Alberts et al. 2008).

In order to attain proper shape a great many nascent proteins depend on "chaperons" (Rutherford and Lindquist 1998; Bergman and Siegal 2003; Taipale et al. 2010; Fig. 5.2). The set of chaperon proteins thus become major regulatory "hubs" that, in different regimes, regulate the cell's crowded protein network by means of fine-tuning (Taipale et al. 2010). In a broader context, not only chaperons but all pre-existing structures and protein assemblages can play formative roles in the environment where a protein molecule is born (e.g., Good et al. 2011). Hence the

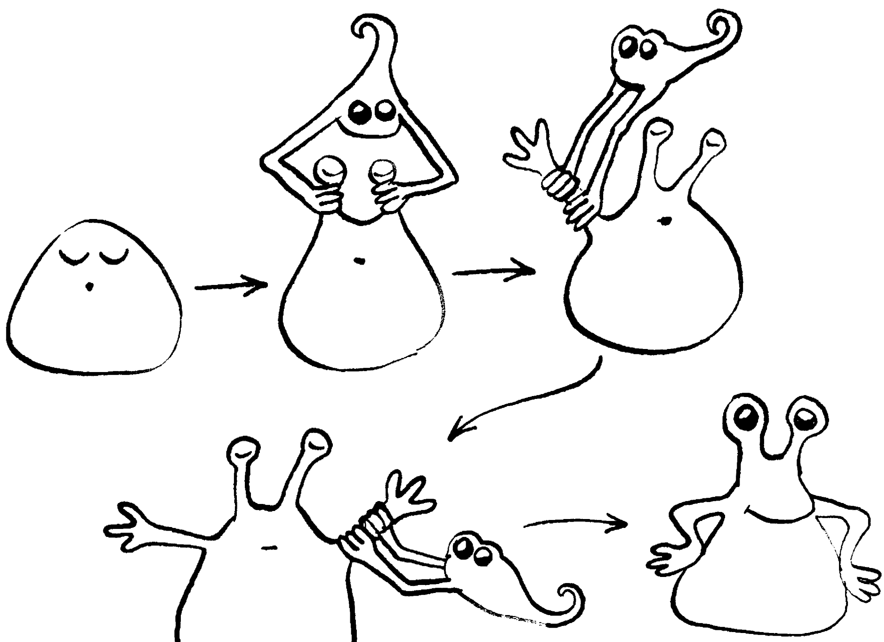


Fig. 5.2 The action of a chaperone on the nascent protein (in many cases, contact with a chaperone is required across the whole lifetime of a given protein)

decision of the context in which the protein is to work is by no means local; it results from the ecosystem of cell "inhabitants." Thus, without any need for central control, proteins function as a distributed meshwork of complex system.

Shape transitions are necessary to protein function. To perform a specific action each must take on a conformation that gives exquisite sensitivity in distinguishing and binding its ligand. On binding, the conformation changes and, by so doing, sets off special operations on or with the ligand. It may, for example, be chemically transformed or transported; a change in conformation may switch a signaling pathway or, perhaps, set off protein-protein binding. The changing conformation can prompt a functional complex to perform a task. The effects of such a change are sketched in Fig. 5.3. During such functions, the protein's performance (speed, efficiency, etc.) may be fine-tuned by the protein ecosystem. While about one tenth of proteins in the cell are bound to "housekeeping" functions (e.g., respiration, food intake, or special syntheses), the others act as a regulatory, information processing network that make subtle responses to whatever happens to the cell.

The function of a protein is *distributed* in that it does not rely on predetermined features alone; it also draws on historical (evolution, ontogeny, given cellular context) and *ad hoc* contingencies (e.g., temperature), or, in short, on the *experience* of

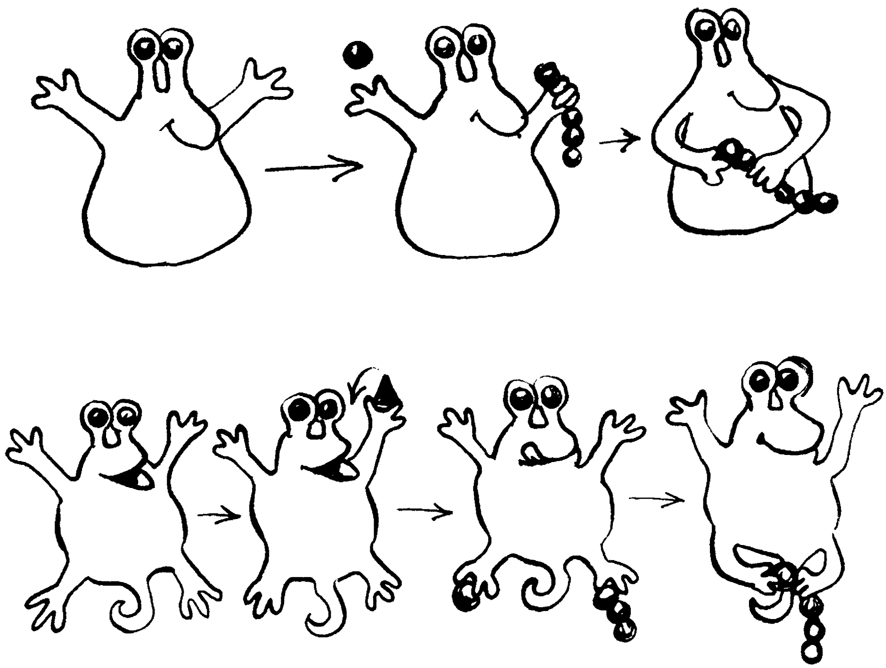


Fig. 5.3 In the *top row* a given protein functions by adding a molecular element to a growing chain. The protein has binding sites to both ligands (the monomer and the chain). Thus, when ligands bind onto specific sites, they induce unifying changes in conformation. In the *lower row* a protein molecule couples with an energy source that enables the inactive conformation to attain the receptive shape required for work (if ligands are available and bound to appropriate sites)

the cell and organism. Such a statement somewhat complicates the straightforward model of evolution described above.

Undoubtedly, evolution draws on random change mutation in the genetic script. As described in every textbook, this leads to alterations in the sequence of protein-coding or regulatory sections of DNA. As a result of change in respective DNA sections, a protein may alter its performance; mutations in the regulatory sequences may also place proteins in new contexts by, for example, altering the timing of ontogenetic gene expression. Changes in the setup of protein network (ecosystem) can have far-reaching consequences for a cell, an individual's appearance (phenotype) and, indeed, for the ecosystem in which it lives. There is, moreover, a second kind of evolutionary change. A whole network of proteins may be induced to change its performance by external factors such as temperature, nutrition, epidemic, that change the appearance and performance of its bearer (some of them outlined in Fig. 5.4). If the whole population is the target of such a change, an unaltered genetic script may nonetheless present a new "attitude towards the world" (see, Matsuda 1987; Hall et al. 2004). Given these two modes of evolution, the network has distributed functions. This is important because, contra the central dogma of biology, this can*not* be traced to inscriptions in genetic code, indeed, it depends on non-local factors that are co-dependent with biochemistry, molecular configurations, function and evolutionary effects. If epigenetic causation (often reversible from the beginning) takes many generations it may even come to be fixed by genetic algorithms (e.g. Waddington 1975; Rutherford and Lindquist 1998). Next, therefore, we turn to how cells develop.

Epigenetics in the Lives of a Cell

Now, we shift our focus from distributed control to consider how a cellular system attunes its current needs by using the 'wording' of genetic texts. We find a sophisticated process that is reminiscent of the subtle use of alternations that "accent" an alphabet's basic letters (e.g. 'a') by marking them as (for example) á, ä, à, â, ã, ǻ, å, ặ, ẵ, etc. While from the point of view of the original Latin such modifications look bizarre, they perform many functions. Even if the differences do not matter at one level (e.g., in e-mails), there are substantial differences at others (e.g., in German, Bar/Bär are different words as are tacher/tâcher in French). In the cell, marks are (reversibly) placed onto DNA or proteins and thus altering the "text" that influences how proteins perform.

Epigenetic use of a diacritical-like processes is far from simple. They ensure, for example, that cells which inherit the same basic 'text' from the zygote can develop into, say, a liver or a brain. As different sets of proteins contribute to the relevant epigenetic processes, organ formation depends on the highlighting and suppression of different parts of genome and/or proteome. There are two key processes in the cell nucleus that help cells (and cell lineages) to remember their spatial and tempo-

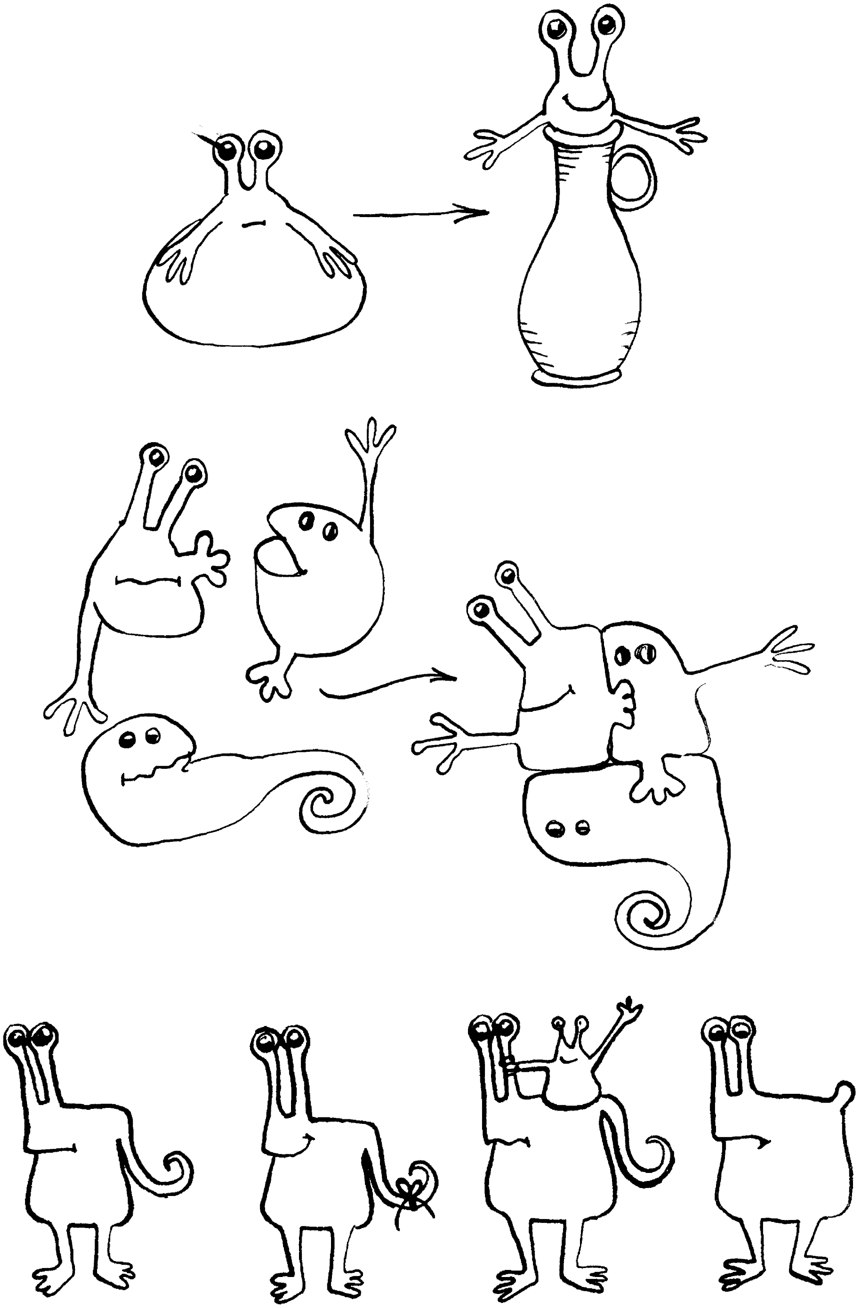


Fig. 5.4 The performing conformation can be also attained by embedding protein into a structural and/or functional context of a specific environment, or can be delicately (or less delicately) and reversibly modified by specific action from its environment

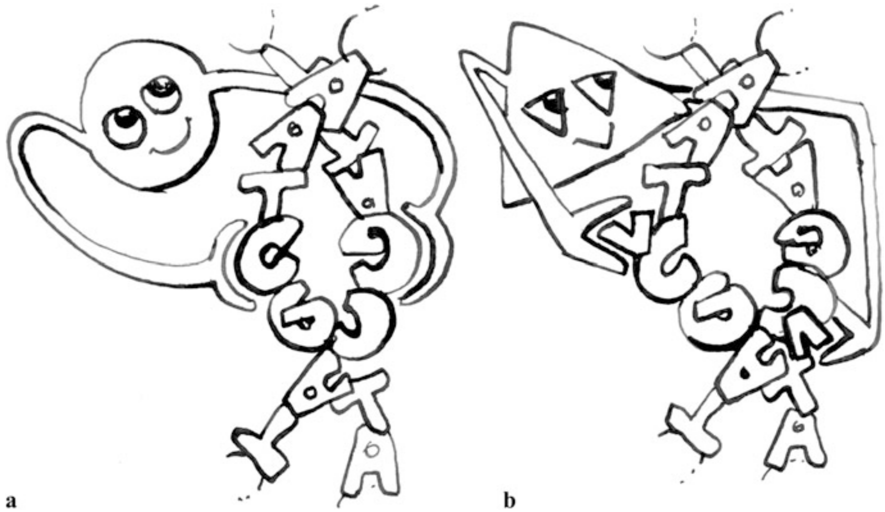


Fig. 5.5 Epigenetic marking: changing some characters affects the overall shape of a section on DNA. If the section AGCTAA represents a *ligand* for a specific regulatory protein (**a**), a modification (to AGČTAA) turns it into *another ligand*; it becomes the target of a protein (**b**). The complex DNA-protein participates in the cell's protein network by influencing its ability to read other parts of the DNA script: the "reading machinery" behaves differently in cases **a** and **b**

ral coordinates.[1] The first of these adds chemical marks (i.e., "diacritics") to DNA molecules. These alter how the script is treated, read and/or understood. The second process uses the organization of the cell's nucleus or scaffolding structures known as nucleosomes. Both processes are tightly interwoven, and deeply influence the cell's orientation and workings.

DNA Markings and Genomic Imprinting

One-dimensional molecules of DNA are often compared to a letter string written in 4 "characters" A, C, G, and T (chemically-nucleotides). On this linear model, the chemical modification of characters resembles human use of diacritics. The commonest of these modifications (methylation) applies to the C character or cytosine in the DNA string. For some mechanisms, nothing has changed (e.g., DNA replication uses the 4 bases); however, for others, the string features a fifth character in the string. Such modifications are reversible in the sense that another battery of enzymes can remove the "diacritics." The method allows methylation to influence the accessibility—and transmissibility—of specific DNA strings. In a reading metaphor, it enhances or hides the text from the performing proteins (see Fig. 5.5).

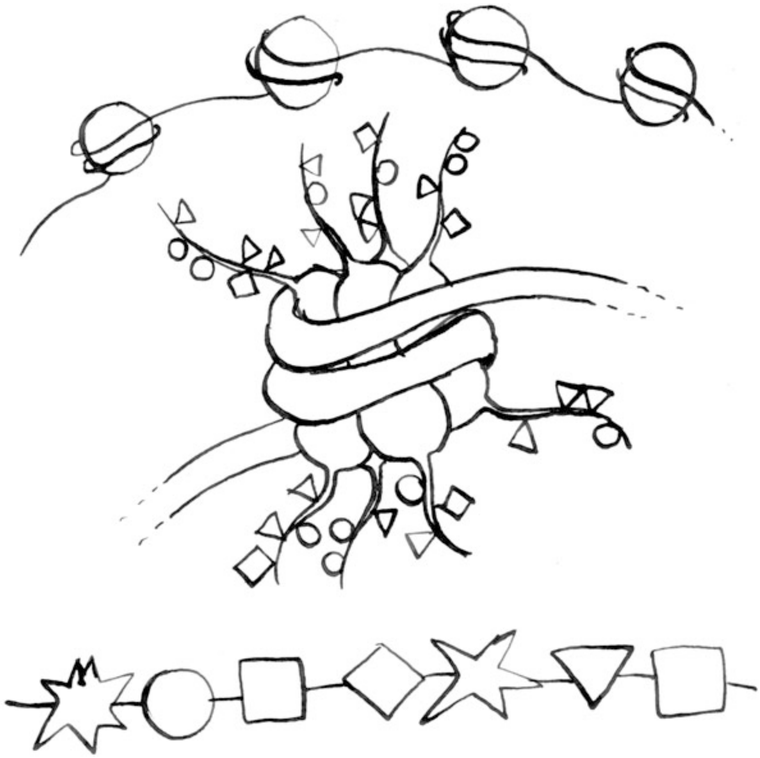
[1] Such processes are especially important in the context of multicellular organisms and their ontogeny. It is important that some of them may outlive even to the next generation, thus transferring the experience of parents.

The reversible process of DNA modification can profoundly influence a cell's internal milieu. This is because it is only by binding proteins to regions of a DNA string that the encoded 'message' can be transmitted to the body-world. Thus, if the functionality of a region is enhanced or hidden, major changes can occur. Such processes therefore function, not only at the level of the cell, but in the organism as a whole. While some epigenetic changes are programmed (as in creating liver cells), others draw on an individual's lived experience. Thus, in identical twins, the pattern of DNA expression is similar early in development. However, across the lifetime, a cascading set of epigenetic effects will draw on processes such as differential DNA methylation.

In other cases, genetic material remembers its maternal or paternal origin. This leads to manifestations in the overall likeness of an individual and is especially well known in so-called genomic imprinting. In mammals, all females are genetic "chimeras" because, in their cells, only one (of two) X chromosomes functions. In a given cell lineage whether this is maternal or paternal is determined at random. If the active chromosome bears a debilitating mutation, the effect cannot be mended in spite of the second (but inactivated) X chromosome has the right gene. Serious mental diseases may develop when maternal/paternal imprinting gets erased or impaired (e.g., Prader-Willi or Angelmann syndromes). In some groups (plants, and perhaps also animals), imprinting enables parents to transmit information to their offspring about the environment they are likely to encounter (e.g., Gilbert and Epel 2009; Allis et al. 2007).

Nucleosomes

DNA strings (billions of "letters"—in mammals) are, in eukaryotic—animal, plant, fungal—cells organized into structures of higher order called *chromatin*: its lowest level of structuration is a "rosary" of *nucleosomes* containing about 147 DNA "characters" wrapped around 8 proteins (doublets of 4 different *histones*, see Fig. 5.6). While stabilizing the strand of DNA, these also enable or deny proteins access to particular sections of genetic material. This depends on functions that are independent of central control. Rather the actions of specific proteins (e.g., methylation, phosphorylation, acetylation), give rise to modifications (and erasures) of histone proteins whose end tails stick out from the nucleosome (e.g., Allis et al. 2007). The modified surface of the nucleosome can thus serve as binding site for proteins that constitute a chromatin ecosystem. Furthermore, such a modification affects all other proteins. It results in a network of interactions that maintains cell differentiation (e.g., as liver cells or neurons) while favouring quick and reversible response to external or internal cues. For example, some genetic material becomes walled up in a given cell lineage or during a developmental stage. By modifying both the DNA and histones, that part of the chromatin acts as an attractor that silences part of the DNA string—possibly thousands of nucleosomes in a row. In other cases, protein assemblies organize regions to produce a given cell lineage. In most cases, even



a

b

c

Fig. 5.6 The nucleosome. **a.** DNA is wrapped around 4 kinds of histone proteins. **b.** Histones are prone to binding by regulatory proteins; epigenetic marking (symbols on protruding "tails") can change the set of proteins that bind to a particular part of a histone. Such a change may switch the whole protein network into a different setting. **c.** Each nucleosome (plus proteins attached to it) thus represents a unique, fine-tuned complex that decides how and when the genetic script at that position is to be read (after Allis et al. 2007)

long-lasting modification may (or should) be reversible in circumstances such as regeneration or, gametogenesis. This view of the cellular ecosystem as akin to reading is shown in the nucleosome pictured in Fig. 5.6.

Elsewhere Markoš and Švorcová (2009) draw an analogy to a natural language that emerges in a natural community of living protein players ("speakers"). This, we argue, cannot be reduced to a fixed code that depends on a program being executed. The parallel is striking: while a histone code can be described in terms of (grammatical) rules, it draws on a dynamical, experience-dependent ecosystem or, simply, the total protein milieu. It is argued that any formal language defined as a set of character strings and determinate operations (Searls 2002) is merely derivative of natural language, that is, it was created by individuals (proteins, cells or humans) who live in the natural world. Developing a consensus on how to read these codes is historical and based on the experience of a community of natural speakers: as Love (2004) suggests, it consists in second-order constructs. Although rules can be described by formal languages, these do not *constitute* natural languages. Just as there are no transcendental laws or rules of human language, biological codes are unlikely to depend on a deeper formal language. Rather, just as in human languaging, biological meaning is extracted by natural 'speakers' who dwell in a historical world of bodily experience.

If the correlation between the DNA script and the shape of the protein is contextual, and experience dependent, then emancipation from the genetic script is likely to go further at "higher," supramolecular levels. Accordingly, we now trace parallels between the interactions of biological systems and the metabolic and symbolic aspects of language and, beyond that, what are usually regarded as different language-systems.

Symbiotic Interactions

In biology, there is often intimate coexistence between two or more lineages of organisms (Sapp 1994, 2003). Such *symbiosis* includes endosymbionts that have been long established within the cells (e.g., the mitochondria or chloroplasts that are viewed as integral to eukaryotic cells), ones living inside other bodies (e.g., bacterial communities in bowels) and the more floating interactions that constitute ecosystems. Symbioses are ubiquitous: they serve the biosphere in that, for example, symbiotic bacteria perform activities that their hosts require. They manage photosynthesis, sulphur metabolism, nitrogen fixation, cellulose digestion, and the production of nutrients (e.g., Hoffmeister and Martin 2003). Symbiosis is thus mainly understood as persistent mutualism or, as "associations between different species from which all participating organisms benefit." Symbiotic interactions are not marginal, academic topic but, rather, resemble the distributed cognitive systems that allow humans to use artifacts and institutions to extend their cognitive powers. In the terms proposed by Douglas (2010): "Plants and animals live in a microbial world. *Their surfaces are colonized by microorganisms* (*bacteria and protists*) *from which they generally derive no substantial harm. Some plants and animals, however, live in specific and coevolved relationships with particular microorganisms, and* these associations have profound impacts on the ecology and evolution *of the taxa involved and, in some instances, also on entire ecosystems. In particular, animal or plants* symbioses *with microorganisms* dominate most terrestrial landscapes, certain coastal environments and the immediate environs of deep-sea hydrothermal vents. [...] *Symbioses are important not just because they are widespread and abundant, but also because the acquisition of* symbiosis can dramatically alter the evolutionary history *of some lineages and change the structure of ecological communities*." (Douglas 2010, pp. 19–23, emphasis added.)

Although symbiosis can be compared with many aspects of human cognition, we focus on its ecological and evolutionary consequences. As an ecological force, symbiosis ensures that species are bound to cohabit. For example, terrestrial plants typically have an intimate symbiotic connection between their roots and fungi. The

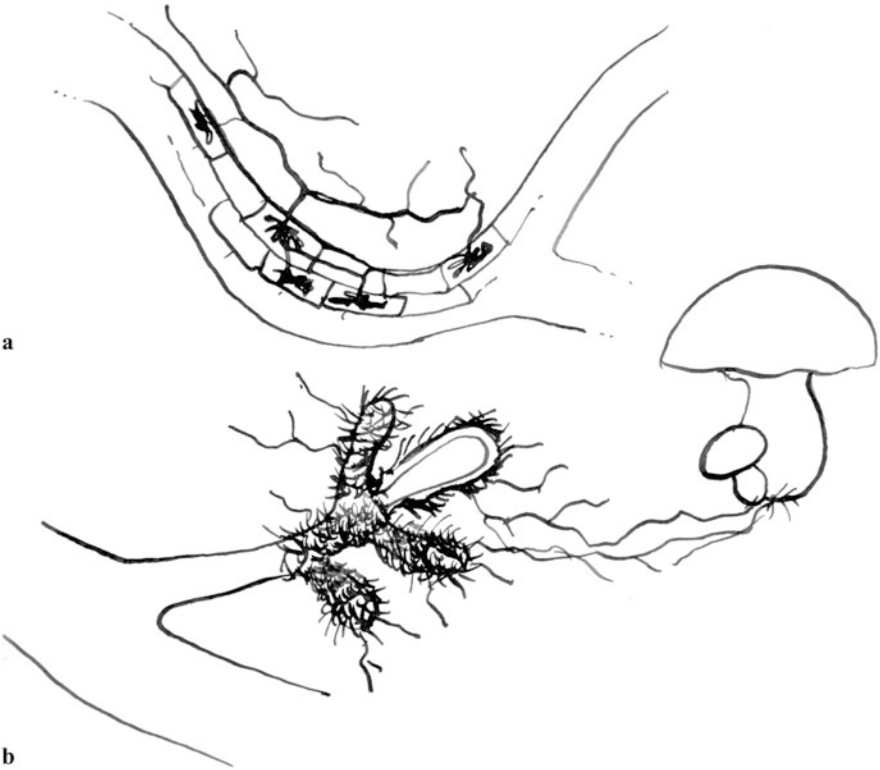


Fig. 5.7 Mycorrhizal symbiosis—tight cohabitation of fungal mycelium with roots of most plants. Two of many possible configurations are shown: **a.** Endomycorrhiza—fine mycelial protuberances invade the plant-cell cytoplasm and create an elaborated network. **b.** Ectomycorrhiza—while also very intimate, hyphae do not invade the interior of cells. The fungus interconnects trees within its reach, i.e. the whole forest may be networked in this way, the network involving many species of plants, fungi, and other organisms like bacteria

most ancient and widespread partnership is *arbuscular mycorrhiza* that dates back circa 460 million years and applies to 250,000 living plant species (Redecker et al. 2000; see Fig. 5.7). Fungi benefit plants by mobilizing nutrients from organic substrates while also delivering water. This is because fungal hyphae are thinner and thus permeate soil better than root hairs. In return, plants subsidize fungi by organic matter.

Symbiosis influences biological evolution profoundly. For example, new lineages of organisms can be engendered by the fusion of previously symbiotically living systems. Symbiogenesis is thought to have given rise to eukaryotic cells that draw on a conglomerate of different bacterial partners (see theory of serial endosymbiosis by Margulis 1993; Margulis and Sagan 2002). Indeed, even those who posit that nature is controlled by something like fixed codes admit that (at least) two kinds of cell organelles—mitochondria and plastids—originated from free-living microbial ancestors (Douglas 2010; Margulis and Fester 1991; Overmann 2006; Paracer and

Ahmadjian 2000; Sapp 1994, 2003). What is remarkable on symbioses is not the fact that different beings, like Russian dolls, share a composite body. Rather, what matters is that, unlike Russian dolls who are indifferent to each other, symbiosis involves mutual understanding between partners who spent even billions of years as separate lineages.

The moral of the story is becoming clear. In biology, wherever we look, we find interactive communities that "somehow" modify what first seems simple. Once we look "below the skin" of a cell, we find an ecosystem of cellular proteins that bend, prune, decorate and tattoo (but also clear away) other proteins: their existence is dependent on a genetic script but their fate depends on the field beyond. The same pattern appears at other levels: although all genes are present in every cell, their expression is distributed through the workings of structures and processes that will put down epigenetic markings. The unpredictability of the outcomes, that is the history of evolution comes to the fore when unrelated lineages enter intimate cohabitations. The same picture applies to ontogenies (i.e., patterning multicellular bodies). Development of a multicellular individual is a fascinating process especially when we trace its historical dimension across lineages and begin to consider what the biosphere has to say about such essentially intimate process.

Ontogeny

Many who discuss evolution echo the central dogma in claiming that *the potential of a species to evolve new traits is constrained by its genome or the set of genes it has available*. For example, Poe writes: "It might be evolutionary advantageous for your progeny to have wings, but it's simply not possible given the genes H. sapiens has to work with" (Poe 2011, p. 8). Whatever the truth of the claim its evolutionary basis cannot be what lies in the genome. Indeed, such a view is the biological counterpart of "written language bias" in linguistics (Linell 2005). Just as written letter strings are sometimes seen as basic, even primary, forms of language, DNA strings can be viewed that way. Function is ascribed to static, reproducible, and rational entities that can be seen and known in totality. Written language bias influenced molecular biology in the 1950s and 1960s (see Markoš and Faltýnek 2011) and, even today, some regard "linear biology" as biological common sense. Just as texts can be reduced to sequences (successions) of letters, DNA conforms to sequences of bases in nucleic acids and proteins. On this view, formal syntax lies 'behind' living phenomena—both language and the likenesses of living bodies. Indeed, the "central dogma" takes the extreme view that information is never ambiguous and flows from a script to the body.

The evidence presented above shows why we reject linear models in biology. First, simple proteins do not derive unequivocal shapes from nucleotides sequences. Second, distributed knowledge contextualizes script by assembling cells whose histories contribute to different lineages and organs. Third, members of different lineages use context to construct a world where cohabitation is widespread. Perhaps, then, we should return to our claim that Ezequiel's meaning cannot be extracted *solely* from a sequence of letters. In denying peace to the wicked (if that was his aim), the likeness (of a message, or of a body) is not a function of a sequence, program, or algorithm. Rather, it draws on a context that belongs to a given lineage, group, organism, and often does so creatively. Members of different species (≡ cultures) treat identical (or very similar) scripts in ways that are quite specific: understanding a text is not a passive crystallization or decoding.

Vertebrates, arthropods, earthworms and even echinoderms have remembered the two-sided symmetry of their ancestors. In the evolution of these *Bilateria*, all species have the same basic body plan (antero-posterior and dorso-ventral axes, left-right symmetry; see Švorcová 2012): differences arise from localized expression of ancient, conservative genes. The body plan is set by embryogenesis long before the appearance of body parts. Since bilaterian phyla have evolved independently for more than 500 million years, it is striking that the basis script remains unchanged. While the genes in each lineage underwent changes in "spelling" as some were duplicated, others deleted or otherwise modified, even unrelated lineages have much in common. For example, deletion of a single gene in the genome of fruit fly can be deleterious or lethal; however, the consequence can be experimentally reversed—by transferring a gene from the genome of a mouse (Gehring 1999). Although proteins coded by such homologous genes differ in many parameters, the message is 'understood': the fly embryo steers the homologue towards a normal developmental pathway. And, of course, "normal" is interpreted as flies (not mice). Thus, if one deletes the gene that initializes eye development in the fly embryo, blind flies will be born. However, a mouse gene restores the development of eyes: those of an insect not a mammal. Thus, a particular protein serves as a tool for establishing a developmental pathway: it does not determine the end product (the eye). Plainly the digital representation of genes (an inscriptional form that may be shared by fruit flies, mice and humans) does not determine how genes work. Rather, this is understood in the "cultural" context of the lineage (species, culture: at the lowest level of description, it depends on an embryo that grows in an ecosystem of interacting proteins—cells and tissues). This complexity allows the same genes to be used in many ways while nonetheless preserving (and transferring) the essentials of the proteins involved. The resulting patterns, ontogenetic outcomes, depend on bodily or lineage memory (see below), not on a linear string that enshrines a memory in a store or depot.

Just as in the Biblical story, the genes are written in an ancient script that is open to non-arbitrary interpretations. Understanding depends on both the individual *and* how the outcome is settled in a given population. The results depend on both situated and non-local factors. To illustrate this matter, one might consider the notorious comparisons between chimp and human genes. While now widely known that their genomes are 98 % "similar," there is debate what such a number means (see, e.g., Marks 2002). Our comparison with reading of book of life can be further elucidated by examples of inscriptions: thus an ancient philosopher's name is rendered as "Aristotelés" in Czech and "Aristotle" in English.[2] Is his message different for both communities? If it is not, as some will argue, this depends on the history of individuals and populations—not spelling.

Examples such as these may appear trivial. However, we should not assume that, in both life and culture, small changes can have large effects. Changing even a fraction of a percent of genetic material can make a difference—especially if the mutation affects a genomic control center (Davidson 2006; Carroll 2005). In presenting our case, we show only that it is naïve to posit the existence of virtual body plans that are attained (and perhaps even foreseen) by a single "keystroke," or a mutation that creates a "hopeful monster."[3]

The High Level of Meaning

We approach the most speculative part of our paper. Rappaport (2010, p. 10) argues: "The survival of any population, animal or human, depends upon social interactions characterized by some minimum degree of orderliness, but orderliness in social systems depends, in turn, upon communication which must meet some minimum standard of reliability if the recipients of the message are to be willing to accept the information they receive as sufficiently reliable to depend upon." We came to similar conclusions earlier when we compared the coherence of members of a biological species to a culture (Markoš 2002). Yet, Rappaport goes even further: his "standard of reliability" lies in rituals shared (albeit not always necessarily respected) by all living members of the community: it is the tie that defines it. Ritual, for him, is "The performance of more or less invariant sequences of formal acts and utterances not entirely encoded by the performers" (p. 24). In other words, it sculpts the "fashion" according to which "we" behave, even if there is no logical necessity to perform exactly in such a way, but it "constructs" the present, as well as the eternity of a given community. "Societies must establish at least some conventions in a manner which protects them from the erosion with which ordinary usage—daily practice—continuously threatens them." (Rappaport 2010, p. 323) "Universal sacred postulates" in rituals serve as such eternal constant that are not to be questioned, not even interpreted in various ways. Yet, they have their evolution across generations. May it be that biological species also constitute such a community kept together by the ritual inherited from the predecessors? Even if rituals seem eternal, they change in subsequent generations as the umwelt or "worldview" of a given lineage shifts in

[2] Versions of written US and UK written English may differ in the spelling in 2 percent of strings. Does this explain the differences between two nations? Remarkably, this line of thinking is pursued by those who seek a genetic Word (in DNA) that is "responsible" for differences in the appearance of the living being (phenotype).

[3] In European history, a single "mutation"—insertion of word *filioque* into the Christian creed (and Son) in the 6th century—is often seen as the main "cause" of schism between Orthodox and Western Christianity.

this or that direction. With a very similar "sacred texts," that is, the genome, we have—after 8 millions years of separation—two cultures of humans and chimps. If the parallel between languaging and life should be fruitful, we should be prepared to think in similar lines. How we look like today is the matter of our genes *and* of the ways how we make use of them in the ever-changing world.

Conclusion

Life cannot be subsumed under physico-chemical principles (even expressible through mathematical notation) because, as Simon (1971) argues, biology and physical science have different objects. Simply said, physical systems lack *meaning*. The fact was first recognized in systems theory and cybernetics (e.g., Bertalanffy 1968); however, no scientific concept of meaning has been developed or needed by the exact and empirical sciences. It is possible, of course, that this is logically impossible or that it just cannot be achieved in quantifiable ways. However, organisms are both ontological and historical: they are products of phylogenetic and evolutionary history. Not only is their multi-scalar nature likely to contribute to the complexity of meaning but this is likely to depend on how relationships use hereditary material to develop over time. As we have seen, this depends on the spatial conformations of DNA molecules and interrelations between them (e.g., DNA-RNA, RNA-protein, protein-protein) that gives the living world has a character of a network or a web of interactions. To grasp the 'core' properties of biological entities, we always need to know about their exact setting. Conversely, it is *far from enough to rely on knowledge of the structure* of their elements. In developing the language metaphor of life (Markoš and Faltýnek 2011; Markoš et al. 2009; Kleisner and Markoš 2005, 2009), we challenge the view that only the lowest level of meaning is accessible to science. Rather, we examine higher levels, where "meaning" gradually becomes applicable to the realm of living. It is our view that this is the most appropriate basis for explaining life and placing it in a coherent system of knowledge that also gives weight to the complexity of human worlds that unfold within a cultural meshwork.

Acknowledgements Supported by the Grant Agency of Czech Republic 13-24275S, by the Charles University Project UNCE 204004 (JS), and by the Charles University Grant 43-25316 (JL). We thank Stephen Cowley for innumerable comments, and for putting the text into tolerable English. Fatima Cvrčková is the cartoons' creator.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2008). *Molecular biology of the cell*. New York: Garland Science.

Allis, C. D., Jenuwein, T., & Reinberg, D. (2007). *Epigenetics*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. The Nobel lecture. *Science*, *181*, 223–230.

Barbieri, M. (2008a). Life is semiosis: the biosemiotic view of nature. *Cosmos and History: The Journal of Natural and Social Philosophy*, *4*, 29–51.

Barbieri, M. (2008b). The code model of semiosis: the first steps toward a scientific biosemiotics. *American Journal of Semiotics*, *24*, 23–37.

Bergman, A., & Siegal, M. L. (2003). Evolutionary capacitance as a general feature of complex gene networks. *Nature*, *424*, 549–552.

Bertalanffy, L. (1968). *General system theory: foundations, development, applications*. New York: George Braziller.

Carroll, S. B. (2005). *Endless forms most improbable. The new science of evo devo and the making of the animal kingdom*. New York: Norton.

Cowley, S. J., & Vallée-Tourangeau, F. (2010). Thinking in action. *AI & Society*, *25*, 469–475.

Davidson, E. H. (2006). *Genomic regulatory systems. Development and evolution*. San Diego: Academic Press.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

Dawkins, R. (1982). *The extended phenotype*. London: Freeman.

Douglas, A. E. (2010). *The symbiotic habit*. Princeton: Princeton University Press.

Eco, U. (1994). *The limits of interpretation*. Bloomington: Indiana University Press.

Gehring, W. (1999). *Master control genes in development and evolution: the homeobox story*. New Haven: Yale University Press.

Gilbert, S. F., & Epel, D. (2009). *Ecological developmental biology: integrating epigenetics, medicine, and evolution*. Sunderland: Sinauer Associates.

Good, M. C., Zalatan, J. G., & Lim, W. A. (2011). Scaffold proteins: hubs for controlling the flow of cellular information. *Science*, *332*, 680–686.

Hall, B. K., Pearson, R. D., & Müller, G. (Eds.) (2004). *Environment, development and evolution. Toward a synthesis*. Cambridge: MIT Press.

Heidegger, M. (1982 [1995]). *The fundamental concepts of metaphysics. World, finitude, solitude*. Bloomington: Indiana University Press.

Ho, M.-W. (1993). *Rainbow and the worm*. Singapore: World Scientific.

Hoffmeister, M., & Martin, W. (2003). Interspecific evolution: microbial symbiosis, endosymbiosis and gene transfer. *Environmental Microbiology*, *5*, 641–649.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: an eternal golden braid*. New York: Basic Books.

Kauffman, S. A. (2000). *Investigations*. Oxford: Oxford University Press.

Kirsh, D. (2010). Thinking with external representations. *AI & Society*, *25*, 441–454.

Kleisner, K., & Markoš, A. (2005). Semetic rings: towards the new concept of mimetic resemblances. *Theory in Biosciences*, *123*, 209–222.

Kleisner, K., & Markoš, A. (2009). Mutual understanding and misunderstanding in biological organisms mediated by self-representational meaning of organisms. *Sign Systems Studies*, *31*, 299–309.

Linell, P. (2005). *The written language bias in linguistics*. London and New York: Routledge.

Love, N. (2004). Cognition and the language myth. *Language Sciences*, *26*, 525–544.

Margulis, L. (1993). *Symbiosis in cell evolution*. New York: Freeman.

Margulis, L. & Fester, R. (Eds.) (1991). *Symbiosis as a source of evolutionary innovation. Speciation and morphogenesis*. Cambridge: MIT Press.

Margulis, L., & Sagan, D. (2002). *Acquiring genomes: a theory of the origins of species*. New York: Basic Books.

Markoš, A. (2002). *Readers of the book of life*. New York: Oxford University Press.

Markoš, A., & Faltýnek, D. (2011). Language metaphors of life. *Biosemiotics*, *4*, 171–200.

Markoš, A., & Švorcová, J. (2009). Recorded vs. organic memory. *Biosemiotics*, *2*, 131–149.

Markoš, A., Grygar, F., Hajnal, L., Kleisner, K., Kratochvíl, Z., & Neubauer, Z. (2009). *Life as its own designer: Darwin's origin and western thought*. Dordrecht: Springer.

Marks, J. (2002). *What it means to be 98 % chimpanzee: apes, people, and their genes*. Berkeley: University of California Press.

Matsuda, R. (1987). *Animal evolution in changing environment with special reference to abnormal metamorphosis*. New York: Wiley.

Monod, J. (1972). *Chance and necessity*. Glasgow: Collins.
Overmann, J. (Ed.) (2006). *Molecular basis of symbiosis*. Berlin: Springer.
Paracer, S., & Ahmadjian, V. (2000). *Symbiosis. An introduction to biological associations*. Oxford: Oxford University Press.
Poe, M. T. (2011). *A history of communications: media and society from the evolution of speech to the Internet*. Cambridge: Cambridge University Press.
Rappaport, R. A. (2010). *Ritual and religion in the making of humanity*. Cambridge: Cambridge University Press.
Redecker, D., Kodner, R., & Graham, L. E. (2000). Glomalean fungi from the Ordovician. *Science*, *289*, 1920–1921.
Rutherford, S. L., & Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature*, *3*(96), 336–342.
Sapp, J. (1994). *Evolution by association: a history of symbiosis*. Oxford: Oxford University Press.
Sapp, J. (2003). *Genesis: the evolution of biology*. New York: Oxford University Press.
Searls, D. B. (2002). The language of genes. *Nature*, *420*, 211–217.
Simon, M. A. (1971). *The matter of life: philosophical problems of biology*. New Haven: Yale University Press.
Steffensen, S. V. (2013). Human interactivity: problem-solving, solution-probing and verbal patterns in the wild. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.
Švorcová, J. (2012). The phylotypic stage as a boundary of modular memory: non mechanistic perspective. *Theory in Biosciences*, *131*, 31–42.
Taipale, M., Jarosz, D. F., & Lindquist, S. (2010). HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nature Reviews. Molecular Cell Biology*, *11*, 515–528.
Thibault, P. J. (2011). First-order languaging dynamics and second-order language: the distributed language view. *Ecological Psychology*, *23*, 210–245.
Waddington, C. H. (1975). *The evolution of an evolutionist*. Edinburgh: Edinburgh University Press.

Chapter 6
The Quick and the Dead: On Temporality and Human Agency

Bent Holshagen Hemmingsen

Abstract This chapter discusses the role of *first personhood*—the experiential dimension—in human agency, and how it informs the agentic trajectories on time-scales from immediate to distal. It is suggested that agency is theoretically divided into two categories, *authorship* and *sponsorship* to establish the differing aims of scientific approaches to agency and a phenomenologically oriented approach focused not on "origin of intent" or "emergence of behavior", but rather on the role of the subjective in maintaining and managing actions and agentic trajectories on different time-scales. The theoretical offset is taken in a theory of temporality, which owes some of its notions to Roy Harris's school of Integrationism. However, it finds the most traction in contested areas between the Integrationist line of thinking, Distributed Cognition, Memory Science, and a few other theoretical lines, which have dissimilar ideas about agency, cognition, and ecology. The neuropsychological case study of CW is examined and discussed to challenge the established views of these theoretical stances on the relation between *first personhood* and agency, and the need to explore how the background state of *sponsorship* facilitates, maintains, or obstructs action is emphasized.

Introduction

Human agency seems a self-evident, yet elusive feature of the human condition. Through the manner in which we commonly make sense of our fellow human beings and their behaviors and actions, we feel compelled to think of them (ourselves included) as capable of acting out their determinations and their intentions to the extent that they are in fact able to make sense of the situations they find themselves in. In a sense we are socially "alive" and temporally "adaptive", or 'quick', rather than 'dead', as the archaic phrase "the quick and the dead" goes. A prerequisite for action is a capacity for distinguishing meaningful contingencies, which support a causal structure turning behavior into evocative gestures of physical, ipseic, interactional, or social "aboutness". The cognitive sciences have found this capacity in a

B.H. Hemmingsen (✉)
Tagensvej 47, 6.-45, 2200 København N, Denmark
e-mail: mult@rapspot.dk

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_6,

theoretical framework purporting neurofunctionality, which is capable of complex computations resulting in situated human action. Whether the origin of action is to be traced to innate systemic biases brought about by ontogenetic or phylogenetic evolutionary calibrations and 'stimuli-response'-like input/output systems are questions without unitary answers.

The more recent Distributed Cognition movement traces agentic gestures in the ecological reality, in which the human individual is entrenched, and the meaningfulness of contributory gestures in supra-individual actions. It is thereby occupied with interactional and cultural dynamics, investigating how ecological events requiring collaborative efforts come into being. This school of thought also claims that in terms of agency that the individual's selfhood is essentially diluted by surrounding factors and co-actors, if not entirely dissolved by dialogicality and contextuality.

Many strong objections towards the lay conception of the responsible, free willed human individual have been made across scientific and academic disciplines. Yet, little energy has been exerted in trying to explain the role of the localized, first person experience—of the experiencing, supposedly capable and responsible individual—and its influence on human agency. It has been dismissed as mere *post hoc* and non-agentic phantasms of the brain. However, one of the problems with trying to link consciousness with what is effectively a vague notion of "free will" *via negativa* is that one is merely trying to trace "origin of intent"—or, as I prefer to call it, *authorship*, and purposely failing to find a connection.

The motivating factors for our actions are only as salient as our creativity in coming up with hypotheses for explaining them. The causal chain always seems to have another link. For present purposes I will use the term *automated* to refer to particle and component motor movements, which make up the sequence of our physical behavior that is not explicitly and willfully brought about. Whether it be speech or bodily movements we do not micromanage our physical behavior, but we do monitor the progression of our intended actions, albeit merely the thematized parts. I propose that this passive monitoring, and what I call *sponsorship* is a *first personhood* capacity, which is required in keeping ourselves engaged in the actions we habitually find ourselves performing.

Furthermore, *first personhood* is riddled with Zeitgebers (Steffensen 2013)—temporality markers—that affords action, interaction, and co-action, as well as giving the human individual a sense of its agentic trajectories. Our *first personhood* temporal capacities make us capable of gauging our agentic conduct on different timescales pertaining to our activities and their dramaturgical composition. However, the most crucial function of human temporality is to help us accommodate to a changing ecological reality, recognizing sameness and change, familiarity and novelty, so we remain current in the very moment we are engaged in, and ward off anachrony. Teilhard de Chardin (1955) wrote of the 'disquiet' as a fundamental disturbance, which rattles our wits and frightens us. As newborn infants we are overwhelmed by the entirety of our new niche in the ecological reality presenting itself to us in a brash and brutal manner, and every moment after our neuronal networks work overtime to catalogue and habitualize to everything we experience, to experientially fasten—'moor'—us to every bit of "salvagable debris" in the raging rivers of Heraclitus. ‘The disquiet’ is paralysis, anachrony, and existential terror, and rarely do we find ourselves entirely at its mercy, but there are those of us who are routinely ‘disquieted’, and they provide fascinating case examples of how the human capacity for *temporality* and agency intersect. One such case is CW, who will be discussed much further later on in the chapter. The following section examines a selection of theoretical stances on temporality, which are pertinent to the view on temporality and agency taken here.

Approaching Temporality Theoretically

Psychological science has traditionally relied on its objects of study being delineable and quantifiable. Recently, the humanities and especially philosophical thinking has had an influence on the rigidness of scientific normativity, especially within certain paradigms of Psychology, for example within Cultural/Narrative Psychology (Bruner 1990) and Discourse Analysis (Jørgensen and Phillips 1999), yet, it goes without saying that neither of these or any other field of science concerned with the biological, cognitive, or social workings of human beings truly account for the dynamics of temporality in such a way that the pursuit of normativity becomes a matter of epistemology first, rather than ontology. Pragmatists (James 1890 [1950]) and phenomenologists alike (e.g., Husserl 1950 [1997]; Merleau-Ponty 1945 [2010]) considered temporal dynamics to be the foundational premises on which things exist and occur, and to be the principles underlying all human perception and *sense-making*. Contemporary thinkers such as Harris (e.g., 1996, 1998) and his branch of Integrational Linguistics (or simply Integrationism) join the ranks of such thinkers, not only by insisting on investigating human characteristics and behavioral conduct from a first person perspective, but certainly also by abandoning normativity, thus, leaving us with a collection of philosophical principles, rather than tools for scientific analysis.

Where Harris, as well as James and the phenomenologists truly differ in their investigation of the human psyche from the science of psychology is in their insistence on taking a first person account, and with that, scrutinizing epistemology and how we come to know and accept “facts”, which are taken for certainties by scientists, no less human than their objects of study. Furthermore, these first person-centric philosophies take temporality to be ingrained in the “fiber” of all being and action, whereas one can easily make the claim that more traditional sciences merely take temporality into account as *ad hoc* stipulation. In traditional science, things are (if not easily, then at least principally) delineable and capable of hypostasis, and problems of emergence, change, evolving, and decay on the one hand, and social and symbolic relations on the other are treated as features and qualifiers, though essentially inconsequential to the production of theoretically ideal types.

Proponents of the Distributed paradigm have a rather different conception of temporality, not focussing on how the dimensionality and directionality of the reality we experience, but instead on the plurality of timescales in which the effects of distributed components collaborate to form aggregates of cognition. There is no emphasis on *first personhood*, but instead a strong emphasis on dialogicality and ecology, and human beings are seen as inescapably entrenched in dynamic systems in which they perform and co-act as contributing actors. The undeclared accord between the Integrationists and the Distributed paradigm is found in their mutual claim that objects of study cannot be informants, when isolating them from their ecological reality that “being and action” occurs inescapably contextualized.

Distributed Cognition and Languaging study behavioral and cognitive dynamics and how their varying timescales interplay, and while temporality plays a role here as well, its objects and subjects of interests belong to a different ontology and warrant a separate epistemological domain from that of the Phenomenologists and Integrationists. Instead of proposing yet another clash of paradigms, I will insist that examining *agency* in the theoretical confines of a study of *temporality* gains far more traction, if one is capable of working at the intersection of two separate epistemological domains, scrutinizing their individual principles and established “facts”, as well as being able to pinpoint areas, where they are able to encourage and strengthen each other’s claims.

Before we move on to some neuropsychological evidence that informs this debate (the case of CW), let us first take a look at Integrationism and the Distributed view, and their attempts with *temporality*-theory, as both schools offer interesting perspectives on human *agency*; often understated, and often at odds with each other. Thereafter, I present and discuss the neuropsychological case of Clive Wearing (CW), which offers descriptions of numerous informative instances of how the dynamics of cognitive traits depicted in either camp’s theory fail to come to fruition in individuals compromised with certain neurological deficiencies. The CW case is particularly interesting because of how time and temporality appear abrupt and disjointed—without order and anchors—to this individual. As he could in several ways be described as “anachronous”, the case provides inspiring challenges to the Integrationist notion of ‘integrating’ and the DCog notion of ‘distributed cognition’, while it tells the tale of an individual, in some way, both experientially unfettered and uninformed by the currents of the physical and social temporal currents flowing and ebbing to and from the ecological reality he used to find himself in.

‘Co-temporality’ and ‘Integrating’

Harris’s school of Integrational Linguistics sets out to dismantle much of the work done in traditional linguistics, disputing such inherent notions as “language is a something”, “words possess build-in meanings”, and that human communication occurs between a sender and a receiver. Because of the deconstructionist and anti-epistemological thrust in Harris’s writing, the reader is often left with sparse tools for scientific and analytical pursuits, but Harris does leave us with a selection of philosophical principles, which set the path for how to approach *agency* and *sense-making*. One of these is the ‘Principle of Co-temporality’ found referenced throughout his works. It dictates that “we are time-bound agents, in language as in all other activities" (Harris 1998, p. 81), and that the "chronological integration of language [and all other activities] with events in our daily lives requires us to suppose that what is said [or done] is immediately relevant to the current situation, unless there is reason to suppose otherwise" (p. 81). The founder of Integrationism, thereby, dismisses the ideal of the study of human communicative conduct *in vacuo*, and similar sentiments can be found in Pablé (2010) who proposes field studies with minimal methodological constraints.

In refuting Austin's (1962) 'illocutionary acts' on the grounds that he finds the distinction between 'performative' and 'constative' to be misleading, Harris (1998) asserts that it is not only ritualized forms of interaction, but all communication that ought to be seen as 'acts' (p. 91). We constantly change the playing field with all our being—sometimes subtly, other times dramatically—whether we set out to do so or not. "Communication invariably takes place in the same time scale and is subject to the same material conditions as all human enterprises. And bringing about the contextually appropriate integration is exactly the task would-be communicators must address" (Harris 1996, p. 14). This is where the rubber meets the road for the Integrationist—that *agency* is an active strive towards the changing conditions of our ecological reality, underlined by such notions as 'semantic indeterminacy' (Harris 1996, p. 84)—that meaning can only be provisionally determined; and 'integrating' (pp. 28–29)—how we attempt to approximate the effervescent, yet seemingly predictable reality.

For Harris, human *agency* is a matter of actively making sense of one's surroundings and actively maintaining the 'integration' between the individual and the world it finds itself in. At times Harris will hint at certain features of human conduct, which he describes as being premised on general underlying capacities, such as memory, but mostly the reader is left with an abridgement of what one could call human cognition. Nonetheless, according to Harris, the world is dynamic and changing, and for the human being to keep up and exert influence, it must be capable of "integrating" into its social environments both by exploiting latent capacities and *sponsored* actions. The Principle of Co-temporality does not imply that experienced reality is in a ceaseless flux, rather it tells us that we are psychologically capable of hypostatizing the real world river of Heraclitus by *anchoring* general-area communicative coordinates ("understanding-for-all-practical-purposes", Linell 1995, p. 181) and thereby *mooring* ourselves "to temporary shores", so to speak. He invites his reader to:

> "Try to imagine ('try' because it is not actually imaginable) that you immediately forgot—permanently—whatever you were doing a second ago, and were unable to anticipate in any way what might happen in the next few seconds or subsequently. If that were the case, you could have no grasp of time. And without that you would not be a human being." (Harris 2009, p. 73)

Thus, foreshadowing the upcoming discussion of the case of CW in the present paper, Harris points to memory as a central feature of the human capacity for temporality, without which we would be incapable of leading the lives that most of us do.

Cognition Gone Wild

Edwin Hutchins's seminal work *Cognition in the Wild* (1995), an anthropological study conducted upon the navy vessel, the U.S.S. *Palau*, was a departure from the classical Internalist view of cognition, purporting cognition as an ensemble of processes occurring within the confines of the individual brains of human organisms. On this seaborne vessel Hutchins discovered cultural ecologies of cognitive workings, which could not be reduced to singular human individuals. Many of the organized routines took place as a result of the complementarity of human individuals performing tasks and utilizing tools, instruments and objects, while being situated in the constraining confines of the ship. In his work, pilotage is described as a process, which is cultured by the evolution of navigation tools, upheld by the combined efforts of the learned crew members working in unison often unmindful of the simultaneous, complementary efforts others exert, while collectively contributing to the accomplishment of various feats.

The implications of Hutchins's writing and the thinking brought about by the Distributed Cognition (DCog) movement in its wake has been a Cognitivist reorientation towards ecological systems in which *agency* arise, stressing the "non-local" (Steffensen 2011), and drawing away from the Internalist and organism-centric approaches, such as the classic Turing and von Neumann inspired views of Cognitivism (Bernsen and Ulbæk 1993). Cognition is seen not only as an aggregation of spatially distributed components, but also temporally distributed components, as the interworking of distributed processes on variable time-scales (Steffensen and Cowley 2010) makes for social and ecological congregations that underlie *living* and ultimately human activity. The cumulative nature of meaning-making is stressed, and thereby another dimension of temporality is addressed in some of these works, as the selective and progressive dynamics of past individual and group behavior grounds all here-and-now *sense-making* and action, though no coherent theoretical outline is attempted for analyzing the role of *temporality* in the light of two key psychological themes, *sense-making* and *agency*. Two areas of the DCog theory very much left without a basic unit of analysis.

The DCog approach is at odds with the Integrationist tradition on several crucial issues, most pertinent to the CW case study is in focus and ontology. Integrationism is undoubtedly first person-centric, championing the experienced dimension of communicative acts and how the individual makes sense of the world and its own conduct. Furthermore, human activity is perceived as active attempts of "integrating", implying a sense of intentionality, which arises out of human *sense-making* in relation to contextual and co-temporal factors. Applying the integrationist metaphor, DCog seems rather to consider human activity as "integrated" (*passive participle*); an inescapable condition rendering the delineation between the human organism and its environment artificial. While both schools are oriented towards ecological thinking, what divides them is an orientation towards or away from the first person experience, and how much *agency* is ascribed to it.

A similar emphasis on ecology over individual can be found in the work of Järvilehto, who maintains that the experienced self is a "systemic relation", and that consciousness is better described as made up of commonality and cooperation, such as common ideas and language, and therefore, we are incapable of going "outside our species and look at ourselves 'objectively'" (2000, p. 91). We are, from this perspective, inescapably entrenched in contingencies emerging out of multitudes of preceding histories shaping, coloring, and guiding our *sense-making* and *agency*. Wegner too promotes an ecological perspective not unlike the ideas in DCog—in his 'transactive memory' (Wegner 1986) and 'co-action' Wegner and Sparrow (2007) theories—which ultimately defines his view of *sense-making* and *agency* as distributed, albeit while maintaining an orientation toward the first person experience. In conjunction with these ideas, DCog theory gains in traction, when we are able to better theorize about the dialogical relationships of agentic influences between *first personhood* and distributed factors. Ironically, the experiential dimension of the human agent in cognitive theory can at times seem almost ethereal, though often entirely immaterial. However, the case of CW does show that there are things to be learned about human psychology from the phenomenological, without it taking away from the central thesis of the DCog program.

The 'Middle World' and Onwards Towards the Inescapable Human Experience

While the proponents of the DCog perspective generally do little to integrate the *first personhood* of the human being into their Distributed Cognitive Systems theoretically, Integrationism only lends us principles, outlines, and a general adversity towards epistemology. Harris's lay-person oriented integrationist perspective shares many of the fundamental reservations against dogma found in Dawkins's notion of the 'Middle World', which alludes to the frame conditions of the experiential dimension of human life. As human beings, on account of the very biological and psychological beings we are, we are simultaneously restricted and enabled in our perceptual and conceptual grasp of the world around us (Dawkins 2006). We are fundamentally unable to perceive or fathom such things as the speed of light on one end of the scale, or a sidereal period at the other end. Neither can we handle microscopic or astronomic sizes very well. We can of course create metaphors and abstractions of such elusive phenomena, processes, and frame conditions with the help of narrative means or measuring devices, through which we are able to quantify and produce ideas of familiar qualities from these esoteric contingencies. Yet, as human beings, we are fettered to species-specific conditions of reality, modulated (of course) by cultured group tendencies and individual experiences (something which Dawkins does not dwell upon too much).

The human 'Middle World' is perceptually and conceptually "located" in between the minute and the grandiose, and while this fact hints at an entirely humbling truth about the human condition—that we will never truly experience the world behind our organs of perception unfiltered—it also underlines an essential feature of our condition: that the human brain potentiates certain types of experiences. *First personhood* 'affords' (Gibson 1986) us contingency-based grounds for *sense-making* and interaction, not to mention the *semiosis* that is the capacity for the active creation of perceived meaningful relations. Meaning is something, which simply is not "out there", independent of the *sense-making* agent.

Authorship, Sponsorship, and Agency

In the study of human *agency* one commonly comes across ideas, which are quite unlike common folk experience, and thereby classifies as contra-intuitive on some level, even to those who propagate such theory. We commonly think of ourselves as capable of making choices for ourselves and taking responsibility of the consequences of our actions. In fact, the 'coming of age' rituals, which are present in most cultures, are rooted in a notion of human agents becoming capable, and to some extent, rational individuals. Yet, at least as early as with Freud, psychology has traced the motivation behind human action beyond the grasp of the rational. Behaviorists saw external stimuli as the primary motivators in the shaping of human behavior, condemning *first personhood* to a "black box" of non-*agency*. Every prominent paradigm thereafter has located the *authorship* of human *agency* beyond the reach of the subjective experience, largely rendering the layfolk perspective utterly erroneous and deceiving in the eyes of theory propagators. The nature vs. nurture debate was at the center of attention throughout the twentieth century, still the "vs. subjective experience" has been conspicuously absent from the discourse.

Recent theorists such as Blackmore (1999) and Järvilehto (2000) see consciousness as a product of the sociality we are entrenched in, albeit pursuing different aims and foci, here *first personhood* is reduced to 'nurture', cloaked by the veils of the "user illusion". Thereby, the experiential basis on which we act and make sense of things is brought about by a social *semiosis*—'memes' according to Blackmore and 'co-operation' in Järvilehto. Both of these theorists promulgate an understanding of consciousness as what one could call "*metaspective*", reducing the role of the experiencer to a mere monological recipient—essentially the singular audience member in the Carthesian Theatre. It would not be improper to apply the term 'automation' to the manner in which human *agency* has been depicted throughout the history of psychology, yet, one would be hard pressed to deny that many of our doings and much of our behavior emerges without detailed micromanagement and explicit rational choice-making. Rarely do we stop to monitor the minute details of the movements put into finding the keys in our pockets and entering them into the keyhole of our front door. We probably do not even separate the acts of returning home from the particulars of the process. Neither is natural speech a product of carefully matching words and intentions of meaning. We are capable of very precise actions, without such micromanaging, but this is not to say that we macromanage our actions in full autonomy either. The Distributed Cognitivists as well as some Dialogism proponents (Linell 2009) see human beings as profoundly entrenched in 'Distributed Cognitive Systems' or 'Dialogical' relations, meaning that no expression of the human organism is ever *in vacuo*; that is, unmotivated by distributed/dialogical contingencies.

Wegner et al. (2004) has sought to trace the cognitive effects, which lead human beings to agentic experiences and attribution of *authorship* to one self. They cite the following seven effects as *authorship indicators*: 'Body and environment orientation cues', 'Direct bodily feedback', 'Direct bodily feed forward', 'Visual and other indirect sensory feedback', 'Social cues', 'Action consequences', and 'Action-relevant thought'. According to this line of thinking, our agentic experience comes from synchronicity of behavior and cues from one or several of these indicators imbuing our sense of own behavior with an experience of *authorship*. What emerges from such pursuits is the question of whether the experience of *authorship* has an agentic effect, and if so, how and to what extent is this so? If there is a correlation between the individual's behavioral self-monitoring, explicitly experienced desires, urges or intentions, and actions, would self-monitoring not be agentic on some level?

Whereas Wegner et al. attempts to uncover the basis for "authorship processing", and to some effect a bias in our experience of *agency* and attribution of *authorship*, a different manner of approaching the issue would be to focus on what difference *sponsorship* or the lack of such does to influence *agency*. Inspired by Harris's (1989, p. 104) claim that "(U)tterances are automatically sponsored by those who utter them, even if they merely repeat what has been said before", the present approach sets out to discover the effects of *sponsorship* as an implicit experiential frame or even mentally thematized on *agency*. When actions are carried out and aligned with intentions (for present purposes), we generally tend to *sponsor* our own actions. However, when our actions fail to meet our intended goals, *sponsorship* start to waiver, and we often end up either distancing ourselves from the outcome of our actions or doubting the *efficacy* (Bandura 2000) of our capacities for carrying out such actions. We aim to make our intentions the conditions of the fluctuating reality in which we find ourselves, and when we fail to do so, our sense of self becomes slightly *anachronous*—in a sense failing to keep up with the currents in the midst of which we find ourselves thrashing about. *Sponsorship* moors the individual to its experience of reality, while the absence of *sponsorship* from the experience of our own actions can make us feel like we are drifting, unable to reach "the shores" that comprise expectations we ourselves and other people set for us.

If much of our behavior is, with lack of a better word, to a large degree *automated*, does this mean that a lack of *sponsorship* influence not only how we come to make sense of our actions, but in fact the execution of actions? If so, it would be difficult to deny *first personhood* a role in human *agency*. The case of CW seems to indicate that when robbed of the continuity, which enables us to update our understanding of self and own actions we are thereby compromised in our capacity for monitoring our own progress with which we gauge whether or not our actions adhere to their intended trajectories.

On Temporality

From a contemporary cognitive psychology perspective, particularly memory research, 'temporality' is investigated as an individual's capacity for moving through time by means of memory recall, prospective memory, and imagination. The focus of studies as those conducted by leading memory scientists such as Tulving (2002), Hassabis and Maguire (2007) emphasizes the two principal qualities of human temporality as being (1) the capacity for traversing the dimension of past experiences and theoretical future scenarios spatially, and (2) the capacity for locating one self in the midst of memories and an evolving "now". I hold a slightly different view of temporality, which I believe sheds new light on the interplay between human *first personhood* and the human cognitive capacities for *sense-making*.

The stance on *temporality* taken here is one, which recognizes the evolving nature of ecological reality, in which the individual finds itself entrenched, though at the same time emphasis is placed on the evolving biology of the human brain, and thereby the dynamic nature of human cognitive functionality. The temporal landscape, wherein *first personhood* unfolds is neither seen as a linear dimension or a raging river of new impressions, which humans are more or less capable of navigating. The spatiality of time is instead seen as dramaturgical in the sense that the relations of salient components—Zeitgebers—are seen as giving rise to experiences of intentionality and causality. The Danish phenomenologist Løgstrup (1976) calls this a "fictitious space", an experiential dimension which is determined by the sequential compositions of its salient parts. In other words, *temporality* in this sense has more to do with relations, patterns, and trajectories informing our sense of understanding of our and other's "being and action" in the on-going ecological reality we find ourselves in.

From this perspective, human *first personhood* is driven towards constancy both by the ceaseless efforts of the cognitive capabilities of the human central nervous system, as well as the conscious efforts in the continued attempts at converting new impressions into recognizable, comprehendible, and meaningful spaces, which human beings grow accustomed to, while learning to navigate the world skilfully. Here the term *temporality* refers to the dramaturgical nature of human *sense-making* capabilities. Essentially "dead" and/or passive surroundings are "quickened" and brought alive. Yet, in this act of "quickening" it is human beings that prove to be 'quick', rather than 'dead', because of how meanings are provisionally instilled into their Umwelt (Hoffmeyer 1993). Human individuals react to changes in the midst of their occurring co-temporally alongside of the meaning that these surroundings afford them.

This notion of *temporality* draws from Harris's principle of 'co-temporality'—that actions are inescapably bound to the temporal context they occur in, and that the temporal context tint the air of the experiencer's occurring experiences. It differs from Harris's scope by recognizing the need for separate, but interwoven epistemologies to describe features of human temporality, which undoubtedly belong to separate ontological domains.

The Curious Case of Clive

The case of the musicologist Clive Wearing (CW) has been well documented in everything from popular science literature (Sacks 2007a), to academic journals (e.g., Wilson and Wearing 1995), a semi-biography narrated by his wife (Wearing 2005), newspaper articles (France 2005; Sacks 2007b), and a few television documentaries (The Mind 1999a, 1999b). Mostly the case subject has been portrayed as an individual, who is completely lost within the currents of time, tumbling back and forth between shallow representations of people, items, and symbols, which individual significance all fail to resonate with him. In Wearing (2005), it is described in some detail, how CW time and time again nearly succeeds in constructing tiny realms of symbolic coherency, yet eventually having that, which he has gained snatched away from him immediately after by the void of his faulty memory. CW is utterly unable bring along with him any new experiences, as he has lost areas of his brain commonly ascribed to recording and managing memory. The singular moments is where he lives, secluded from most of his past, his recent-past, and his future, thus, lacking his basis for temporal integration.

CW was a musicologist and conductor, who had worked for the BBC among other companies, and had before being struck by illness been working a seven day week regularly, trying to squeeze in his own pursuits, such as conducting and intense studies of chorus and chamber music into his already busy schedule as a producer. He had been especially interested in researching minute details of prominent past choral performances, which had taken place at famous royal European courts during the Renaissance. His attention to detail was described as impeccable, and as his wife writes about the scraps of notes, photographs from specific past concert locations: “Our photographs were all marred, indecipherable, but the contents were held in CW’s head. The rooms of our flat were lined with manuscripts, jottings, paraphernalia that meant nothing to anyone but CW, but through his knowledge, his understanding, there came from these items sublime music.” (Wearing 2005, p. 50).

In March 1985 tragedy struck for the Wearings, as CW was struck by the herpes simplex encephalitis virus, which would eventually render his hippocampus and parts of his frontal and temporal lobe regions largely into dead nerve tissue due to necrosis (Hansen and Knudsen 1983). Consequently, he would struggle to make sense of the details of his experiences, rather than tirelessly pursuing them with his usual gusto. In his own words, he described every conscious moment as if he had just woken up from a long slumber with no recollection of the past moment The duration of his effective “present” moment was estimated to be somewhere between seven to thirteen seconds, sometimes up to as much as half a minute without much sense of continuity—utterly *in medias res*.

The Fading Familiarity of Things

In her book *Forever Today*, Deborah Wearing gives an emotional account of her life with CW, his neurological condition—primarily anterograde amnesia along with a severe degree of retrograde amnesia—and her descriptions of the consequences of its effects in their everyday life. Anterograde amnesia is characterized by the lack of ability to form new episodic, autobiographical, and often semantic memory as well, and it commonly follows from lesions or necrosis occurring in the hippocampus regions of the brain. The inability to form new memories hinders the accommodation to a changing reality, which essentially leaves the suffering individual in a constant state of anachrony, or in other words without the means to anchor new meaning and exploit it to navigate the ceaselessly evolving reality one constantly finds oneself in.

Retrograde amnesia is functionally very different, as it mostly signifies the selective loss especially of episodic, autobiographical and semantic memories—past experiences and knowledge about self and the world—while the individual usually remains functionally capable in terms of forming new memories. Whereas primarily anterograde amnesic individuals in a sense move further and further away from the world they recognize, retrograde amnesic individuals merely have lesser or larger voids in their memories in terms of prior experiences and knowledge, with which they used to feel comfortable and familiar.

Additionally, anterograde amnesic individuals tend to lose many of the layers of the social life in which they have previously been embedded soon after the onset of the condition. Their everyday lives tend to become significantly more simplified in terms of relational behavior, as more complex or temporally distal goals take a backseat to more immediate matters. Familial and societal expectations lessen, and there is a significant diminishing of personal authority concerning decision-making and setting one's own everyday priorities. This is arguably because the amnesic individual is incapable of "putting things in larger perspective", especially where distant either past or forthcoming preconditions factor in.

CW is no exception. From the depiction we get from his wife, he is rarely without some sort of supervision, and since he is incapable of retaining his thoughts and immediate experiences for more than a mere moment, he tends to either get easily fixated or side-tracked. This being an obvious handicap in terms of the type of social *agency*, which is defined by an individual's capacity for maneuvering social reality and ability to alter the stipulations into which he defines his social selfhood, CW is never completely devoid of that *strive* of intentionality towards fellow beings, objects of desire, and psychological states of mind that we commonly ascribe to lucid human beings. Even without the sensitivity for much of the complex symbolism that requires one to retain and cross-reference past knowledge, prior events, and future causations, CW seems undeterred and anything but apathetic in trying to figure out the moments he "awakens" to.

CW has provided many recorded first-hand accounts of his immediate experience; for example "I wasn't conscious, I have no consciousness at all! Consciousness has to involve ME!!!" (Wearing 2005, p. 204). CW's depiction of his condition is interesting for at least these two reasons. First, his memory deficiency corrupts his sense of continuity in terms of his *self*. He is aware that he is watching a video recording of himself and that the entries in his diary was written by his own hand. Yet, he completely refuses to acknowledge this prior self as having been conscious at the time, as he sees and understands himself as being awake for the very first time in a long while, at the very instance he witness himself in action or re-reads his notes (The Mind 1999a). In his notebook one finds entry after entry crossed out and erased for the simple reason that they are not, according to CW, either factual or actual recordings of his "authentic self". In other words, he does not experience any sense of *authorship* in relation to these actions, and he hesitates to *sponsor* them. Second, his sense of *self* is to some extent crystallized, as the implicit updating of his register of self-related knowledge occurs at an extremely gradual rate, and thereby he remains far less dialogically agile in terms of attempting to remain current and relevant in his interactions with others.

The apparent anachrony with which CW enters social time does frequently frustrate and bewilder him. Yet, he seems to find loopholes in his condition or rather he exploits some of the capacities, which most people take for granted as part of our human repertoire in his attempt to compensate for his deficiencies. Here Hutchins's sense of 'Distributed' cognition and the idea of 'external representations' (Kirsch 2010) seems ideal as conceptual frames for understanding how CW's agency is integrated in an ecology of sociality and physicality, which helps him anchor meaning and moor himself to an evolving reality,[1] when his memory fails him. This case also shows how the capacity for the distribution of complex cognitive tasks relies on localized cognition, without which, the 'external representations' fail to "represent", that is fail to support the individual in his integrating behavior.

Furthermore, the CW case points to the role of first personhood and the importance of sponsorship for agency. Jones (2010) makes the case that morals inform our actions, and essentially that ecologizing agency is akin to silencing the individual as an informant of its experience of taking stances and making rational choices, and reducing the individual to a mere "cog" in the " 'distributed' machine". CW is often stumped in his actions, because he fails to make sense of the situation in which he finds himself, thereby finding himself incapable of *sponsoring* what he perceives to be disjointed experiences of own behavior.

First personhood is often overlooked as a problematic informant of the psychological workings of human beings both *in situ* and *in vacuo*, yet, the CW case can be seen as an indicator of the importance of *sponsorship* for not only "directed", but also "automated" behavior. That is, not only the behavior we are aware of and are monitoring, but also the behavior, which is not apparent to our own selves, or which we do not need to monitor closely. The miniscule changes in positions and movements of one's arm, when reaching for a glass of water to take a sip are fully "automated" in the sense that no other part of *first personhood* enters into this action than the very experience of intention, and inexplicitly, the background state of *sponsorship*. One could say that *sponsorship* is the absence of alertness or apprehension, or perhaps even closer to what is intended by the term could be described as the nonappearance of any experienced 'disquiet' threatening to disrupt the unity of selfhood and action.

[1] Similar to the DCog perspective and pertinent to the topic of memory, Wegner (1986) provides the theoretical concept of 'transactive memory' to describe the social distribution of memory tasks.

In the Absence of Coherency and a Sense of Trajectory

When it comes to applying explanatory models to the interactional stratum of the organism-environment system (to borrow Järvilehto's phrase) Clark cautions against taking a purely dynamicist stance—that is to emphasize the role of interactional dynamics without taking ontologies of localizable systems into account. According to Clark (2008) the individual organism is a salient component of what DCog proponents would term the 'Distributed Cognitive system'. Clark (2008) writes that "[a]s soon as we embrace the notion of the brain as the principal (though not the *only*) seat of information-processing activity, we are already seeing it as fundamentally different from, say, the flow of a river or the activity of a volcano" (p. 25), thus setting the tone for his separation of 'knowledge-based' (embodied) and 'physical-causal' (distributed) systems. Drawing on his idea of 'continual reciprocal continuation' (Clark 2008, p. 24)—that one system is continually affecting and simultaneously being affected by some other system—Clark sets out to account for the temporality of this condition, and he does so by invoking the two aforementioned separate, but interworking ontological and temporal dimensions. Yet, *first personhood* is a separate, but integrated ontological and temporal dimension, which is unaccounted for in Clark's writing on the matter.

In the documentary *The Mind: Life Without Memory* (1999a), CW is depicted as an individual without the ability to anchor his past and without a clear sense of his agentic trajectories. Furthermore, he is described as highly emotional, most likely because he has very few moments of serenity and few co-temporal "sanctuaries", wherein he can find solace from the constant alteration of the ecological reality in relation to his experiential positioning. What Clark's dualism fails to describe is how emotions help us form a dramaturgical sense of coherency and trajectory in the experienced sense of self and the world, both of which occur in the realm of *first personhood*. When CW's attention is drawn towards the trivial, such as when he is suddenly pre-occupied with whether or not his wife's blouse was exactly the same color a minute ago that it is now (Wilson and Wearing 1995, p. 17), he shows an indication of being 'disquieted'. Much the same, when he is found repeatedly covering and uncovering his one hand with the other, revealing and concealing a piece of chocolate, which he claims is of a different kind than the one he believes is supposed to be found on the palm of his hand (Wearing 2005).

CW does not lack the motivation or will to act or respond to his surroundings, but he does lack the means to familiarize himself with and accommodate to a changing reality cumulatively. To him, temporal dynamics reveal themselves in a far more imposing manner quite different than to most of all other human individuals. To those of us with functioning memory, the temporality is implied in our experience of the world, yet it is rarely very apparent. Our experience of the on-going progression of reality mostly appears to be coherent, rather than jagged and disjointed, and this is so because of our functioning cognitive capabilities, tidying the elements, which enter temporally into the perceived and experienced. When there does appear to be inconsistencies, we react in a similar manner to CW with emotional arousal, cautioned by the appearance of uncertainties. This state of 'unquiet' urges on a demand for order and coherency in the world we find ourselves in, and in our physical, social, and psychological positionings. Without coherency, we cannot *sponsor* the relation between our experiencing *self* and our experienced *selfhood* or our *selfhood* and its relation to other elements in our ecological reality. For CW his actions are often *in medias res*, accompanied by a shattered sense of their agentic trajectory; presumable, he rarely has a clear sense of what he sets out to do, and exactly where he imagines that his actions will take him.

James (1890 [1950], pp. 449–450) wrote of emotions that the causal chain goes as following: we perceive the threat, we run, and as we run we experience fear, reversing the chain of events usually attributed to the experience of emotions. This has later been supported by the work of Damasio (1994 [2006]) and the experimental work of Levenson et al. (1990). It gives further credence to what Tinbergen (1963) termed a 'proximate mechanism', that is, a local cognitive functionality preceding the experiential dimension of *first personhood*. However, the claim being made here is that human agency arise from the dialogical interplay of the neurofunctional domain and *first personhood*, and that the case of CW points to how memory capacities facilitate an evaluative dimension, wherein more complex problems can be gauged with the nuances of the total strata of felt human emotion and *sense-making* dramaturgy. Furthermore, *first personhood* evaluation not only respond to symbols and affect, but does so on a rather different timescale than what occurs in the neurofunctional domain (Tulving 2007), although it goes without saying that nothing occurs in human experience that does not also involve neurons firing. The case of CW does provide us with problems worth investigating, if we are to understand the role of *first personhood* in human agency. His experiential dimension is claustrophobically tight, and only fairly simple chains of events can unfold within the span of his experienced "now". Such a limitation restricts the scope of his agency, indicating that human beings need not only memory, but 'explicit memory' (Zola and Squire 2000 [2005]).

Further Perspectives on First Personhood, Memory and Human Agency

Tulving (2002) describes the case of K.C., an individual with a similar neurological condition as CW, suffering from "global episodic amnesia" (pp. 316–317). He writes the following:

> "[K]nowledge of time in and of itself does not allow [K.C.] to remember events as having happened at a particular time. It is necessary but not sufficient. Something else is needed, and this something else—the awareness of time in which one's experiences are recorded—seems to be missing from K.C.'s neurocognitive profile. Thus he exhibits a dissociation between knowing time and experiencing time, a dissociation between knowing the facts of the world and remembering past experiences." (p. 317)

Elsewhere, Rosenbaum et al. (2009) tie anterograde amnesia, specifically in the case of KC, to a deficiency in capacities for imagining fictitious or future scenarios, as well as “a general inability to generate details and/or bind them into coherent representation” (p. 2182). The work of Hassabis and Maguire (2007) support this finding, and lend credence to the position of the ‘constructive memory’ paradigm, which central tenet holds that remembering or recalling past experiences is a constructive, rather than a reconstructive act. Phrased differently, it holds that acts of remembering are functionally similar to acts of imagining in terms of both of them being creative processes, and essentially that they are not acts of re-creating or recalling verbatim depictions of past events. This paradigm goes back at least as far as to Bartlett (1932 [1955]), and there are parallels to this line of thinking in the more contemporary empirical work of the neurochemist Nader (2007).

The purpose of drawing attention to these studies and this research is to reiterate and emphasize the central thesis that serves as the theme of this chapter—that regardless of where we end up pursuing the actual *authorship* of human individuals’ agency, there is still the nagging problem of the relationship between the cognition that occurs and persists on the level of neurofunctionality and the narrative, or rather the dramaturgical semiosis, which ontologically belongs to a separate domain entirely. It is here that we find *sponsorship*, grounded in an individual’s capacity for making sense of its ecological reality, while undergoing a polyphony of various emotions; “moving” the individual as he positions his experiencing *self* in relation to his experienced sense of *selfhood* (Mead 1934; Hermans 2001). The *sponsorship* occurs as a relation between the *self* and the *selfhood*. The less the *selfhood* stands out, the more the experiencing *self* is pre-occupied with its pursuits. A *selfhood* that is frequently present at the center stage of the individual’s attention, is a presentation of self indicating a recurring issue, obstructing the individual’s strive towards its desires.

A sense of trajectory, though not necessarily accompanied by a detailed plan for navigating present conditions, seems sufficient for maintaining the agentic thrust towards goals or end states in one’s actions. James (1890 [1950], p. 627) describes the following scenario:

> “A road we walk back over, hoping to find at each step an object we have dropped, seems to us longer than when we walked over it the other way. A space we measure by pacing appears longer than one traverse with no thought of its length. And in general an amount of space attended to in itself leaves with us more impressions of spaciousness than one of which we only note the content.”

A familiar situation, which might imply that when we lose sight of our trajectory, or are torn out of what Csikszentmihalyi (1997) calls the ‘flow’, we are in a sense *disquieted* by the lack of implicit sense of certainty. Losing sight of trajectory elongates the experience of time, because the uncertainty of the elements in our ecological reality leaps to our attention, but fail to be evocative of productive meaning. We encounter this modulation of temporal dramaturgy with an estrangement from our own current pursuits, because of how our emotive states rush the predicament to the forefront of *first personhood* experience. The sense of urgency of that we undergo never feels greater than at times where the absence of automaticity and unhindered ongoing action is prevalent. This is where *sponsorship* waivers, and where we start to question our agentic *efficacy*, even if just for the moment.

Løgstrup's (1976) 'fictitious space' occurs at the intersection between processes of memory and ongoing experience. It is an ontological domain, tingling with explicated objects and topics, which enact the relations we have with outer and inner world conditions on a stage and with the props of obscured and implied semantic and dramaturgical directionality. Emotions are the soil, on which *sponsorship* grows, just as they are the recontextualizing obstructions, reframing the focus of attention, and underlining the current, specific hardship proving to be a challenge. They motivate and demoralize action—responding, not to any certain kind of stimuli, but instead—evaluating contextualized effects of our ecological reality based on the imprints left on our neuronal structures by prior experiences. Response occurs on many separate and intersecting levels, and to reduce cognition to a matter of input/output or stimuli/response is to completely overlook the plurality of microprocesses, as well as the roles of the separate systemic macro-evaluations occurring interminably, and dialogically influencing each other in the process of *sense-making* and producing behavior.

Concluding Remarks

Integrationist ideas concerning temporality and agency bring about a markedly different and admittedly first person-centric take on psychological areas of interest, one of them being agency. As a lay-person oriented school of thought its proponents has concerned themselves with what the *first personhood* informs them of, in respect to the human condition. In this regard, the integrationist shares a common interest with James, whose studies were heavily informed by 'introspection' as a method of investigating the dynamics of "mental life". Yet, with these approaches, there is a danger of disregarding the dynamics of other ontological domains, which interplay with *first personhood* to bring about *agency* and thereby the human agent, whose actions are a product of a plurality of biological, neurological, and social *authorships*, as well as the evaluative systems, which are subtly and sometimes explicitly expressed in the *sponsorship* or lack of that the human individual undergoes and experiences. I have departed slightly from Harris's original use of the term, although I feel that my version adheres to integrationist principles in a way that brings about a new manner of rethinking psychological theory, without succumbing to "surrogationist" attempts at scientifizing the school's original stride and purpose.

The integrationist writings imply an autonomic volition, which beckons further investigation and explanation, but the basic human strive for things exterior of its own physical extremities cannot be properly accounted for without accepting the need for separate epistemologies. Some of which, has little to do with *first personhood* all together. Consciousness does not "do things" or "perform actions", but it is rather that "fictitious space", wherein one's ecological reality is experienced, emotions are felt, and objects of interest are thematized. It is here that we see urges and intentions come to light, and their meaningfulness reflects back dialogically on other human subsystems, which in turn modulate neuronal patterns and eventually physical behavior along the lines of the agentic trajectory we feel that we are setting for ourselves.

Consciousness is therefore neither 'non-agentic' (Skinner 1974), or an amalgamation of pure social permeation (Blackmore 1999; Järvilehto 2000). It is 'metaspective', but certainly agentic. Not in the sense of *authorship*, but in terms of *sponsorship*. The case of CW teaches us, from a third person perspective, what it means to constantly experience the lack of *sponsorship* and undergo one *disquieting* after the other. Essentially his memory and his emotions are not complimentary of each other or congruent, and in his attempts to accommodate to evolving reality, the absence of coherency produces a constant set of problems in need for him to attend to, in a never ending string of futile attempts to restore sense to the ecological reality of his *first personhood*. Yet, the incoherency is motivating as well. It is agentic in the sense that it produces discomfort or even terror in the individual, urging the need to return sense and congruence to one's on-going experience. If consciousness was indeed an "afterthought" or even "non-agentic", we would not react so strongly, whenever we experience that some feature of our *first personhood*—serenity, the absence of pain, functioning perception, the capacity for remembering selectively, a sense of purpose etc.—has been *disquieted*. Instead we would go about our business, none the wiser, which is clearly not the case. *Sponsorship* makes us adaptable on a different level, accommodating to changing realities. Principally, this is the difference between the 'quick' and the 'dead'.

References

Austin, J. L. (1962). *How to do things with words*. Oxford: Clarendon Press.

Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science*, *9*, 75–78.

Bartlett, F. C. (1932 [1955]). *Remembering: a study in experimental and social psychology*. Cambridge: Cambridge University Press.

Bernsen, N. O., & Ulbæk, I. (1993). *Naturlig og kunstig intelligens: introduktion til kognitionsforskningen*. København: Nyt Nordisk Forlag Arnold Busck.

Blackmore, S. (1999). *The Meme machine*. New York: Oxford University Press.

Bruner, J. (1990). *Acts of meaning*. Cambridge: Harvard University Press.

Clark, A. (2008). *Supersizing the mind: embodiment, action, and cognitive extension*. New York: Oxford University Press.

Csikszentmihalyi, M. (1997). *Finding flow: the psychology of engagement in everyday life*. New York: Basic Books.

Damasio, A. (1994 [2006]). *Descartes' error*. London: Vintage.

Dawkins, R. (2006). *The god delusion*, London: Black Swan.

France, L. (2005). The death of yesterday. *The Observer, Sunday, 23 January 2005*. http://www.guardian.co.uk/books/2005/jan/23/biography.features3. Accessed 16th of July 2010.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale: Erlbaum Associates, Inc.

Hansen, K., & Knudsen, G. M. (1983). Nervesystemets infektioner. In O. B. Paulson, F. Gjerris, & P. S. Sørensen (Eds.), *Klinisk neurologi og neurokirugi* (pp. 437–456). Copenhagen: FADL's Forlag.

Harris, R. (1989). How does writing restructure thought? *Language & Communication*, *9*, 99–106.

Harris, R. (1996). *Signs, languages and communication*. London: Routledge.

Harris, R. (1998). *Introduction to integrational linguistics*. Oxford: Pergamon Press.

Harris, R. (2009). *Integrational notes & papers 2006–2008*. Bedfordshire: Bright Pen.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, *11*, 299–306.

Hermans, H. J. M. (2001). The dialogical self: toward a theory of personal and cultural positioning. *Culture & Psychology*, *7*, 243–281.

Hoffmeyer, J. (1993). *En snegl på vejen—Betydningens naturhistorie*. Copenhagen: OMverden/ Rosinante.

Husserl, E. (1950 [1997]). *Fænomenologiens idé—fem forelæsninger*. Copenhagen: Hans Reitzels Forlag.

Hutchins, E. (1995). *Cognition in the wild*. London: MIT Press.

James, W. (1890 [1950]). *The principles of psychology* (Vols. 1 & 2). New York: Dover.

Järvilehto, T. (2000). Consciousness as co-operation. *Advances in Mind-Body Medicine*, *16*, 89–92.

Jones, P. (2010). You want a piece of me? Paying your dues and getting your due in a distributed world. *AI & Society*, *25*, 455–464.

Jørgensen, M. W., & Phillips, L. (1999). *Diskursanalyse som teori og metode*. Frederiksberg: Roskilde Universitetsforlag.

Kirsh, D. (2010). Thinking with external representations. *AI & Society*, *25*, 441–454.

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, *27*, 363–384.

Linell, P. (1995). Trouble with mutualities: towards a dialogical theory of misunderstanding and miscommunication. In I. Marková, C. Graumann, & K. Foppa (Eds.), *Mutualities in dialogue*, Cambridge: Cambridge University Press.

Linell, P. (2009). *Rethinking language, mind, and world dialogically: interactional and contextual theories of human sense-making*. Charlotte: IAP.

Løgstrup, K. E. (1976). *Vidde og prægnans. Metafysik 1. Sprogfilosofiske betragtninger*. Copenhagen: Gyldendalske Boghandel, Nordisk Forlag A/S.

Mead, G. H. (1934). *On social psychology (selected papers, edited and with an introduction by Anselm Strauss)*. Chicago: The University of Chicago.

Merleau-Ponty, M. (1945 [2010]). *Phenomenology of perception*. Cornwall: Routledge.

Nader, K. (2007). A single standard for memory; the case for reconsolidation. *Debates in Neuroscience*, *1*, 2–16.

Pablé, A. (2010). Hvordan forholder integrationismen sig til den virkelige verden? *Spindet, Årgang*, *10*, 16–21.

Rosenbaum, R. S., Gilboa, A., Levine, B., Winocur, G., & Moscovitch, M. (2009). Amnesia as an impairment of detail generation and binding: evidence from personal, fictional, and semantic narratives in K.C. *Neuropsychologia*, *47*, 2181–2187.

Sacks, O. (2007a). *Musicophilia: tales of music and the brain*. Toronto: Random House.

Sacks, O. (2007b). The Abyss: music and amnesia. *The New Yorker, September 24, 2007*. http://www.newyorker.com/reporting/2007/09/24/070924fa_fact_sacks.

Skinner, B. F. (1974). *About behaviorism*. London: Jonathan Cape.

Steffensen, S. V. (2011). Beyond mind: an extended ecology. In S. J. Cowley (Ed.), *Distributed language* (pp. 185–210). Philadelphia: Benjaminss.

Steffensen, S. V. (2013). Human interactivity: problem-solving, solution-probing and verbal patterns in the wild. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Steffensen, S. V., & Cowley, S. J. (2010). Signifying bodies and health: a non-local aftermath. In S. J. Cowley, J. C. Major, S. V. Steffensen, & A. Dinis (Eds.), *Signifying bodies: biosemiosis, interaction and health* (pp 331–355). Braga: The Faculty of Philosophy of Braga.

Teilhard de Chardin, P. (1955). *The phenomenon of man*. New York: Fount Paperbacks.

The Mind (1999a). *The mind episode 10: life without memory: the case of Clive Wearing, Part 1*. Colorado: Colorado State University. (2 Videoclips). Accessed 18th of July 2010.

The Mind (1999b). *The mind episode 11: clive wearing, Part 2: living without memory*. Colorado: Colorado State University (4 Videoclips). Accessed 18th of July 2010.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, *20*, 410–433.

Tulving, E. (2002). Chronesthesia: conscious awareness of subjective time. In D. T. Stuss & R. C. Knight (Eds.), *Principles of frontal lobe functions* (pp. 311–325). New York: Oxford University Press.

Tulving, E. (2007). Coding and representation: searching for a home in the brain. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Coding and representation: searching for a home in the Brain* (pp. 65–68). New York: Oxford University Press.

Wearing, D. (2005). *Forever today*. London: Doubleday.

Wegner, D. M. (1986). Transactive memory: a contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). New York: Springer.

Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: experiencing control over the movement of others. *Journal of Personality and Social Psychology*, *86*, 838–848.

Wegner, D., & Sparrow, B. (2007). The puzzle of coaction. In D. Ross, D. Spurrett, H. Kincaid, & G. L. Stephens (Eds.), *Distributed cognition and the will: individual volition and social context* (pp. 17–37). Cambridge: MIT Press.

Wilson, B. A., & Wearing, D. (1995). Prisoner of consciousness a state of just awakening following Herpes Simplex Encephalitis. In R. Campbell & M. A. Conway (Eds.), *Broken memories: case studies in memory impairment* (pp. 14–28). Oxford: Blackwell.

Zola, S. M., & Squire, L. R. (2000 [2005]). The medial temporal lobe and the hippocampus. In E. Tulving & F. I. M. Clark (Eds.), *The Oxford handbook of memory* (pp. 485–500). New York: Oxford University Press.

Chapter 7
You Want a Piece of Me? Paying Your Dues and Getting Your Due in a Distributed World

Peter E. Jones

> *'What you think cognition is and what you believe is part of the architecture of cognition depends on what you imagine to be typical or important cognitive tasks and what you think a person is' (Hutchins, Cognition in the wild, 1995, p. 316).*
>
> *'There are issues of pride, passion and politics involved, not to mention intelligence and imagination, and ultimately—perhaps initially and primarily—moral responsibility as well. And they are involved not merely as contributory causes or consequences but as substantive questions concerning how—if at all—A communicates with B' (Harris, Signs, language and communication, 1996, p. 2).*

Abstract The chapter offers a critical reflection, inspired by the insights of integrational linguistics, on the conception of thinking and action within the distributed cognition approach of Edwin Hutchins. Counterposing a fictional account of a mutiny at sea to Hutchins's observational study of navigation on board the Palau, the paper argues that the ethical fabric of communication and action with its 'first person' perspective must not be overlooked in our haste to appeal to 'culture' as an alternative to the internalist, computer metaphor of thinking. The paper accepts Hutchins's own critique of the 'meaning in the message' illusion but goes beyond this critique to argue for a view of communication, thinking and action as creative, ethically charged and morally accountable acts of engagement.

In this paper I wanted to bring issues of agency and personal responsibility into the discussion of thinking and action and to do this via a reflection on a number of communicative acts, some actual, some fictional. This ethical slant on things first occurred to me as I tried to articulate a response to some of the other chapters in this volume. While activities of various kinds—practical, communicative and

P.E. Jones (✉)
Department of Humanities, Sheffield Hallam University, City Campus, Howard Street, Sheffield, S1 1WB, UK
e-mail: p.e.jones@shu.ac.uk

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_7,

cognitive—were being investigated, questions about the rights or wrongs of the actions involved or about who should take responsibility for them were not on the table. It then struck me that this conspicuous absence of ethical considerations might tell us something about the value of the approaches to language, thinking and action which influence and guide our research agendas. Of particular interest was the question of how this ethical theme might sit within the socially- and culturally-oriented considerations of the 'distributed cognition' tradition, exemplified by Hutchins's (1995) pioneering study of navigation aboard the *Palau*. Furthermore, I saw an opportunity to magnify the ethical theme by exploring another, and sharply contrasting, case study of communication in a military setting. This case study is, I confess, a fictional one. But it has the merit of posing in dramatic style the ethical questions I wanted to focus on here.

In fact, putting your disinterested observer's standpoint to one side and asking questions about who is responsible for an action and whether or not you agree with or support it, whether you would have done the same thing in their shoes, etc., is a good way to probe the limits—arguably, the limitations—of the varied intellectual endeavours usually referred to as 'cognitive science'. For, after all, in everyday life (outside the laboratory or research paper) we are constantly having to make judgements about the rights and wrongs of particular actions, about who is responsible for something and how much of the credit (or blame) we should assign to them, and about whether we should do something, get involved, etc. Such judgements may be very difficult, or even quite arbitrary at times, but we cannot avoid either making them or being on the receiving end of them ourselves. Making such ethically informed judgements is itself, of course, an activity subject to ethical scrutiny for which we might have to take (or attempt to shirk) responsibility. So there is no way to put our actions and feelings in principle beyond ethical judgement and no action or feeling that might, under certain circumstances, suddenly acquire an (un)ethical value where it had none before. And this applies to linguistic or communicative practices just as much as anything else (see Jones 2007).

Questions of agency and ethical responsibility often show up most strikingly when things go either spectacularly wrong or spectacularly right. When, apparently against all odds, something goes right, we often celebrate the skill, experience, judgement, intelligence, heroism, and moral rectitude of an individual, often to the detriment of other members of the team that made the successful outcome possible. The skilful landing of an ailing passenger plane on the Hudson River in New York is a case in point. As *The Guardian* put it: 'Captain Chesley "Sully" Sullenberger kept a low profile yesterday, but that didn't stop him being called the "hero of the Hudson" for his calm, deft response in the face of calamity, after he safely landed his passenger plane in the river off New York on Thursday' (Saturday, 17th January 2009: 'Captain Cool Ex air force pilot an expert on aviation safety'). The article goes on to mention that 'New York's governor, David Paterson, said an anonymous person had offered to donate $10,000 toward building a statue to the pilot'.

In fact, we might rather cynically argue that any reference to 'his calm, deft response in the face of calamity' is either speculation or wishful thinking for all we actually know about what went on. In that connection it is interesting that *The*

Guardian's accompanying feature entitled 'Crash course: Tips for passengers—and pilots', we see that kind of embellishment from ignorance exposed: 'Someone suggested the pilot chose the Hudson River as the plane would land near the rescue services. I doubt he had time to factor that into his calculations. He had to choose between landing on water or the skyscrapers of Manhattan'.

Taking the point to extremes, you could imagine a comedy sketch where the hapless Sully, not having a clue what to do and panicking hysterically in the cockpit, accidentally knocks against some obscure little switch marked 'Emergency Hudson Landing' and then cowers on the floor praying hard till he realizes that the plane has landed safely. Recovering his composure and adjusting his tie he walks out onto the flight deck looking 'absolutely immaculate', 'like David Niven in an airplane uniform'.

When things go wrong, on the other hand, we see a tendency to fling the guilt around pretty indiscriminately. When two 10-year old boys abducted and killed a toddler in Bootle, Merseyside in 1993 everybody and everything was fair game, from the boys themselves, their family upbringing, the estate where they lived, the video nasties they were alleged to have watched, their schooling (or lack of it), to Thatcherism and the collapse of morality in contemporary society. Blame was well and truly distributed, although in the end only the boys themselves were tried, convicted and locked up.

I offer these examples in the interests of making two pretty obvious points. On the one hand, individuals do not do what they do, say what they say or think what they think in a social or cultural vacuum. Their actions both presuppose an unfathomably complex history of interaction with other people in particular contexts and also connect them afresh to other people (ultimately, everybody) directly or indirectly. In that sense, the responsibility for any action, good or bad, is always 'distributed' between all those who contributed to it, wittingly or unwittingly. But, on the other hand, whatever the difficulties and dilemmas involved, we generally cannot avoid distinguishing between all kinds of contributory (or mitigating) factors or conditions and actual personal responsibility (guilt or innocence) and degrees thereof. Why so? Because actions are not performed by abstractions (like 'culture') or impersonal 'systems' but by real people. And often—sometimes despite ourselves—we have to step up and put ourselves on the line. So reflecting on behaviour in these terms helps us to become aware of an essential dimension of human thought and action which is often left, at best, at the theoretical margins.

By the same token, such questions also provide the occasion for a revealing contrast between the ('third person') 'analysis' we might present of human activities in our professional capacity as psychologists or linguists or ethnographers and the kinds of ('first person') value judgements and reactions we might offer in our capacity as, for want of a better term, ordinary human beings. In Jones (2007) I consider the value of this more down to earth approach for exposing problems with 'Critical Discourse Analysis'. In this chapter, with the help of a little bit of Hollywood glamour, I apply a similar approach as a way of helping unravelling the knotty problem of the relations between communication, thought and action as they might be seen from a 'distributed' perspective.[1]

Ghosts in the Machinery

In spite of the lessons life continually throws at us, some academic disciplines, as we all know, actually make a virtue of having no relevance to the slings and arrows of everyday fortune. Orthodox linguistic theory is a very good example of one such. The more formal approaches, Chomskyan linguistic theory being the best example, have vigorously resisted any encroachment of what we might see as human values on their theoretical territory. Instead, we are offered a 'scientific' perspective on language in which real living and breathing individuals are effaced by abstract—even 'universal'—rules, principles, parameters, codes and the like.[2] Roy Harris has spoken eloquently of the ethical cleansing of language and communication that has resulted from the procedures and perspectives of orthodox linguistic theory: 'Thus a form of discourse about language is created which serves either to disengage language from human motives and intentions, or to disguise the extent and nature of that engagement. Through this discourse language is presented as being in itself neutral, a mere communal instrument or facility. Ethical, political and aesthetic judgements are passed only on particular "uses" of language by particular individuals or groups in particular circumstances' (1987, p. 162).

The passage is taken from Harris's (1987) book, *The language machine*. As the title may imply, the book is a severe critique of the attempt to picture communicative and cognitive activity in terms of an impersonal mechanical or (more recently) computational process at work within the individual mind. The 'myth of the language machine', as Harris puts it, 'absolves us from our day-to-day duties as language makers, and blankets out for us all awkward questions concerning the exercise of authority through language' (1987, p. 162). The result, as Harris argues, is 'the removal of language from the domain of social morality altogether' (1987, p. 162). Against which Harris insists that 'any premeditated act of communication is itself a moral act' (Harris 2004, p. 728).[3]

[1]It is interesting to note that other socio-culturally oriented approaches to the understanding of human activity and thinking are also wrestling with the problem of how to give the individual his or her due within their analyses. Anna Stetsenko, for example, argues that A.N. Leont'ev's 'Activity Theory' places too great an emphasis on the socially determined character of activity, thereby 'positing society above the individual and seeing the latter as produced by, subordinate to, and molded by reality, and especially society, at the expense of emphasising individual agency—the ability to produce, create, and make a difference in social practices' (2005, p. 78). See Halverson (2002) for a comparison of 'Activity Theory' and 'distributed cognition'.

[2]See Love (2007) for a discussion of the view of language as a 'digital code'.

[3]Bert Hodges, too, has insisted on an ecological position according to which 'realizing values is central to language' (Hodges 2007, p. 585) and has explored the implications of that position for linguistics, language learning and psychology.

The morally-purged view of language which orthodox linguistics has given to the world has also been very influential in psychology. Indeed, the Chomskyan version of the 'language machine' myth has provided a rationale and a model for the very development of 'cognitive science'. However, Hutchins (1995) is a valiant attempt to break free from the dead hand of that tradition. His own forthright critique of the 'mind as machine' perspective in the history of cognitive science chimes with that of Harris. However, Harris, but not Hutchins, also sees and foregrounds the ethical implications of the analytical reduction of language to sets of mental symbols manipulated according to fixed rules. For that reason, as we shall see, the influence of the 'myth of the language machine' is detectable, if muted and often contradicted, even in Hutchins's remarkable 'distributed' exploration of collaborative navigational practices.

An immediate response to these points might invoke the need for 'inter-disciplinarity'. The linguist, for example, in defence of intellectual territoriality or scientific 'idealization', might concede that his or her theoretical constructs need to be supplemented by somebody else's to get a fuller, rounder picture of actual communicative activity as carried out by real individuals. But this argument obliges us to accept that the phenomena identified and treated by the self-standing disciplines have validity to start with, something that we have good reason to be sceptical about.

In the case of linguistics, for example, we could challenge the very procedure via which a conception of language as a computational machinery is developed in the first place. What is it that underpins the mindset of the theoretician who finds an impersonal system at the foundation of linguistic activity? And what kind of psychology or sociology will you have to make do with to complement *that* take on language and communication? Two or three wrongs are unlikely to make a right. Putting the humanity back into the study of human linguistic activity (and everything connected with it), then, would seem to require not so much *inter-disciplinarity* but *contra-disciplinarity*—going against the theoretical grain, not with it. And I hope that this paper may demonstrate some of the virtues of such a contrarian stance. (In passing, it is interesting to note the flirtation with 'machine' discourse to be found amongst ethnomethodologists and Conversation Analysts, otherwise well known for their opposition both to a view of language as an internal, individual capacity and to the reification of social interaction represented by such concepts as 'the culture' and 'the language', Silverman 1998, pp. 48–50.)[4]

Silverman concedes that Harvey Sacks 'seems to imply a very mechanical model' of conversational interaction 'in certain parts of his lectures' (1998, p. 65). For example, Sacks describes interactions "as being spewed out by machinery, the machinery being what we're trying to find" (1998, pp. 65–66). Furthermore, despite the evident differences in subject matter and methodology between CA and linguistics, Sacks was influenced by Chomsky and considered linguistics as 'the most advanced social science' (Sacks in Silverman 1998, p. 50).

[4]On the relationship between ethnomethodology and Conversation Analysis see Schegloff's 'Introduction' in Sacks (1995).

Now, of course there is more than one way of avoiding or obscuring the ethical dimension of human communicative or cognitive behaviour. One way, as already noted, is to simply exclude such considerations on theoretical grounds. Another way is to choose for investigation practices or events which appear to incite no moral qualms, either because they are harmless tasks artificially constructed by researchers or because, being trivial and mundane activities, they are seemingly uncontroversial. The standard laboratory or experimental psychological task offers an example of the first type (and some of the tasks reported in this volume are of this kind). Hutchins (1995) gives us an example of the second type. His account of the routine navigational work of the crew of the *Palau* does not appear to offer us anything ethically contentious, even though this is the American navy in action.

But then imagine your reaction to finding out that the ship had been involved in a politically sensitive mission, one that we are strongly opposed to on political or ethical grounds (e.g., the invasion of Iraq). The question of responsibility—personal and collective—for this abominable action might then, quite rightly, become *the* issue before all others, as happened in the case of the torture and abuses committed at Abu Ghraib. And arguments, or excuses, to do with the weight of 'culture', 'ideology' or binding military discipline would probably not cut much ice either when apportioning blame.

So, as Hutchins himself says in the passage at the beginning of this chapter, what we think a person is has a lot to do with the kinds of behaviour we think are important or typical. In that connection it is interesting and, possibly, significant, that Hutchins tells us that his 'initial assumption about work in military settings was that behaviours are explicitly described and that people act more or less as automatons' (1995, p. 225). A strange assumption to make in the first place, you might think, although it is a relief that he immediately follows up with the comment: 'It should be apparent by now that this is far from the case' (1995, p. 225). And if it had not been apparent, one might add, then Hutchins's book would have been a very bizarre and troubling read as an account of human action. Instead, Hutchins gives us a remarkable description and analysis of team effort and performance. More to the point, it delivers, sometimes directly and sometimes in passing or implicitly, a series of stunning critiques of conventional positions in the study of mind, language and communication.

All the same, you will search in vain in Hutchins's account for anything significant on the ethical dimension of human activity. This is, at least partly, understandable and excusable on the basis of Hutchins's attempt to demonstrate a distinction between, as he puts it, 'the cognitive properties of the sociocultural system and the cognitive properties of a person who is manipulating the elements of that system' (1995, p. 362). But it is also undoubtedly a function of his choice of subject matter, a very particular 'universe of social arrangements' (1995, p. 202) defined by naval role and rank. Here Hutchins finds 'a tissue of human relationships in which *individual watchstanders consent to have their behaviour constrained by others, who are themselves constrained by the meaningful states of representational technologies*' (1995, p. 202, my emphasis). And it is this state of affairs that Hutchins clearly considers to be both typical and important for a more general reflection on the nature of human thinking and action.

As impressive as Hutchins's study of the *Palau* is, however, I believe I can go one better in the interests of demonstrating the 'ineluctable' moral dimension of communication (Harris 2004, p. 728). What I have to offer is nothing less than a case study of a mutiny on board an American nuclear submarine. This time we're not dealing with a situation where individuals 'consent to have their behaviour constrained by others'. Far from it. Constraint 'by the meaningful states of representational technologies' is sometimes hard to find as well.

The Crimson Tide

The situation I am talking about revolves around one of the most interesting lessons in the perils of communication that Hollywood has ever devised, namely the movie *Crimson Tide*, by the director Tony Scott.[5] Denzel Washington and Gene Hackman face off as, respectively, XO (Executive Officer) Hunter and Captain Ramsey of the nuclear submarine *Alabama*. In Russia, rebels have gained control of their country's nuclear capability and are readying missiles for launch. 'There's trouble in Russia, so they called us', Ramsey announces to his assembled crew as they prepare to depart. At sea, the *Alabama* receives an order to prepare its own nuclear missiles for launch within the hour to destroy the Russian missile silos (and most of Russia with them). The order comes in the form of an 'Emergency Action Message' ('EAM'), duly authenticated according to a strict matching procedure on board. As the clock ticks and the *Alabama*'s missiles are close to launch state, another EAM starts to come through but the radio equipment fails before a complete message (with the necessary authentication code) can be received. XO Hunter seizes the print-out from the radio operators and takes it directly to the Captain:

XO: Got the EAM. What do you think?
Captain: I think there's nothing on this.
XO: Yes, sir. It got cut off during the attack.
Captain: Then it's meaningless.
XO: Sir, this is an EAM pertaining to nuclear missile—
Captain: No, Mr Hunter. That is a message fragment.
XO: Because it got cut off during the attack, sir. The message could mean anything. It could be a message to abort, it could be a message to—
Captain: It could be a fake Russian transmission.
XO: Which is exactly why we need to confirm, sir.

But the Captain is in no mood to listen and walks away. The XO attempts to engage once more and the conversation picks up again:

[5]The movie released in 1995 by Hollywood Pictures. The dialogue is taken, with some of my own alterations, from the website: http://www.imdb.com/title/tt0112740/quotes.

Captain: We have orders in hand. Those orders are to make a pre-emptive launch. Every second that we lose increases the chances that by the time our missiles arrive, their silos could be empty because they've flown their birds and struck us first.

XO: Yes sir.

Captain: You know as well as I do that any launch order received without authentication is no order at all.

XO: Yes sir.

Captain: That's our number one rule.

XO: [*tries to get a word in*] National Mil—

Captain: And that rule is the basis for the scenario we've trained on, time and time again. It's a rule we follow without exception.

XO: Captain, National Military Command Centre knows what sector we're in. They have satellites looking down on us to see if our birds are aloft and if they're not, then they give our orders to somebody else. That's why we maintain more than one sub, it's what they call 'redundancy'.

Captain: I know about redundancy, Mr Hunter.

XO: All I'm saying—[*Ramsey walks off*]

XO: [*follows Ramsey, lowers his voice so that the rest of the crew can't hear*] All I'm saying, Captain, is that we have backup. Now it's our duty not to launch until we can confirm.

Captain: You're presuming we have other submarines out there, ready to launch. Well, as Captain, I must assume our submarines could've been taken out by other Akulas. We can play these games all night, Mr Hunter, but I don't have the luxury of your presumptions.

XO: Sir—

Captain: Mr Hunter, we have rules that are not open to interpretation, personal intuition, gut feelings, hairs on the back of your neck, little devils or angels sitting on your shoulder. We're all very well aware of what our orders are and what those orders mean. They come down from our Commander-in-Chief. They contain no ambiguity.

XO: Captain—

Captain: Mr Hunter. I've made a decision. I'm Captain of this boat. NOW SHUT THE FUCK UP!

Unable to agree on a course of action, and each man unwilling to back down, the two key protagonists ignite a violent civil war onboard which pits two groups of officers and men against each other for control of the nuclear launch codes. As the dialogue makes clear, at the heart of the conflict is a something whose very identification and description is a matter for contestation. For XO Hunter, this something is a message, an EAM; for Captain Ramsey, it is a 'message fragment'. For Hunter, it 'could mean anything'. For Ramsey, it is 'meaningless'. Ramsey justifies his categorization of the something with reference to strict military discipline and the set rules for identifying and handling messages. By the criteria he cites it cannot be counted as a message at all. Hunter, on the other hand, begins from a different perspective. His argument is not about the 'meaning' of the message according to the rules, either of the navy or of ‘English’. As a close look at the sheet of paper on the screen shows, there is ‘nothing on it’ of any use or relevance in the wording and no possibility of authenticating its status or provenance. The ‘rule we follow without exception’, in Ramsey’s words, rules it out of account. The film throws this clash of readings in our faces and pulls us viscerally into the conflict: Whose side are we on? Who is right and who is wrong? What would we do in this situation?

The Meaning in the Message

The *Crimson Tide* dramatizes a conflict playing out at various levels. On one level, this is personal. Hunter and Ramsey do not like one another. In the heat of the moment of confusion provoked by the aborted signal this personal animosity, a clash of personalities, fuels their intransigence. But it is also a difference of what we might call ‘personal philosophy’, as the film implies in a prior episode. Around the officers’ mess table, in a moment of relaxation and informal conversation, Ramsey asks Hunter if he thinks it was right to drop the atom bomb on Hiroshima. Hunter says it was but weighs his words and conveys no enthusiasm for the decision. The Captain takes this up:

Captain: You do qualify your remarks. If someone asked me if we should bomb Japan, a simple ‘yes, by all means, sir, drop that fucker, twice!’ I don’t mean to suggest that you’re indecisive, Mr Hunter. Not at all. Just, uh, complicated. ’course, that’s the way the Navy wants you. Me, they wanted simple.

XO: Well, you certainly fooled them, sir.

Captain: [*chuckles*] Be careful there, Mr Hunter. It’s all I’ve got to rely on, being a simple-minded son of a bitch. Rickover gave me my command, a checklist, a target and a button to push. All I gotta know is how to push it. They tell me when. They seem to want you to know why.

XO: I would hope they’d want us all to know why, sir.

This discussion of philosophical, or cultural, differences sets the scene for the later confrontation over the problematic EAM. In trying to explore this confrontation, we might draw on the insightful discussion of communication in Hutchins (1995). Ramsey, as Hutchins might say, subscribes to a communicational approach according to which the meaning is very definitely ‘in the message’ (1995, p. 238). Things mean what they mean by virtue of their possession of a form and content prescribed in advance. In this particular case, the ‘message fragment’ does not measure up against anything recognizable. It simply lacks the requisite form and content; what form it has is illegitimate, and it is semantically empty. But, as Hutchins explains, this particular communicational philosophy is based on an illusion: ‘Meanings can only even be imagined to be in the messages when the environment about which communication is performed is very stable and there are very strong constraints on the expectations. In many endeavours, creating and maintaining the illusion that meanings reside in messages requires that a great deal of effort be put into controlling the environment in which communication takes place. Meanings seem to be in the messages only when the structures with which the message must be brought into coordination are already reliably in place and taken for granted. The illusion of meaning in the message is a hard-won social and cultural accomplishment' (Hutchins 1995, pp. 238–239).

Indeed, Ramsey himself spells out how the illusion is created—through training 'time and time again' on particular scenarios and drumming in a set of expectations and responses. It is not, then, that the message contains an order which must be obeyed. Rather, it is the fact that the crew have been trained to conform in their expectations of and responses to particular phenomena that gives the artefact its status and meaning as an order. They seem to be doing what things tell them to do. But things don't tell you what to do. In effect, the training process is one of suppressing or anaesthetizing the individual's moral sense. In this, the members of the crew are actually, as Hutchins argues, *consenting*. They are *consenting to submit to authority and, thereby, to suspend their ethical responses and judgements*. What we read 'in the message' is nothing but our own conformity to a particular 'universe of social arrangements' (Hutchins). If the rules are 'not open to interpretation' (Ramsey) it is only because the people who create them, read them and send them have been trained—or coerced—to think and act within strict parameters and to execute orders without question (i.e. without regard for the consequences of those orders).[6]

The power of the *Crimson Tide* is that, in a single moment of verbal conflict, the 'meaning in the message' illusion is shattered. The smoke and mirrors are cleared away, the smelling salts are administered and the ethical fabric of human interaction is suddenly nakedly exposed.

XO Hunter, for his part, appeals to a very different communicational ethic. The Captain, Hunter can concede, definitely has a point about the interpretation of the radio transmission. But, unfortunately, the course of action he proposes as a consequence is going to result in nuclear holocaust. For Hunter, the meaning is not 'in the message' but in the whole situation they—and the rest of humanity—are facing. While prepared to justify his own position by appeal to procedure and authority where he can, Hunter sees that the very fact of the arrival of a 'meaningless' message fragment is meaningful in itself, right here and now, whatever the rule book says and however powerful is the impulse to obey a direct order. When the stakes are this high, we have the right, the duty, to withdraw our 'consent to be constrained by' the behaviour of others; we have the right to refuse the constraint of 'meaningful states of representational technologies'. But then we also need the moral courage to exercise that right and to take responsibility for what follows from it.

Hunter's refusal to support the Captain polarizes the ship. His stance throws all positions and attitudes into ethical relief. The invisible seam of conformity to command becomes a jagged edge of armed confrontation. For some crew members, the legalistic protocol of command procedure—which Hunter is quick to exploit in his favour—offers a reason, or at least a pretext, for supporting the XO against the Captain. For others, the Captain's course of action is the right one, whatever the rule

[6]On the creativity of conformity see Hodges (2007).

book says. But as the parties violently engage it becomes impossible not to take sides. The 'universe of social arrangements' itself can now be judged as to its ethical value. Simple unthinking compliance with orders is now suddenly revealed and evaluated as an ethical stance in itself. The social and moral glue holding the familiar meanings and messages together has dissolved and everything is up for grabs. 'They tell me when' is not good enough any more. Now we need to know why.

Messages and Meaning: Distribution and Integration

Hutchins's critique of the 'meaning in the message' illusion strikes at the main pillars of conventional theorising about language according to which language is a complex code, that is, a system of linguistic forms (e.g., words or morphemes) along with rules for combining them plus a set of meanings to be expressed or realized by these forms. This is basically the position in theoretical linguistics that Harris (1995) refers to as 'segregational' on the grounds that it assumes from the outset that something called 'language' exists as a self-contained and coherent object for scientific investigation independently of the actual practices of real people in real situations. In contrast, Harris's (1995) 'integrational' approach proceeds from the observation that human behaviour offers no grounds in principle for a distinction between the linguistic and the non-linguistic since human sign-making activities (of which what we call 'language' is one such activity) are integrated aspects of the conduct of concrete individuals. As Harris puts it: 'Episodes of communication are episodes in the lives of particular people at particular times and places. Signs are products of such episodes. They do not exist independently in some quasi-Platonic realm, or psychologically as subsistent "mental representations" (e.g., stored in the long-term memory of the individual, or a hypothesized collectivity of individuals)' (2009, p. 70).

From this 'integrationist' perspective, then, Hutchins has gone some way but could have gone further.[7] In the case of the *Alabama*, something is clear which is perhaps not so clear on the *Palau*: it is not merely a question of what meaning (if any) we give to a message but whether there is a message at all. What is at stake is the freedom, ability, or right of one individual to see things differently, to see things that 'the culture', or the 'universe of social arrangements' hasn't prepared for or bargained for. In the case of the *Alabama*, the conflict erupts over Hunter's decision that there is a message—a meaning to be derived—where others do not, will not or cannot see one. The conflict between Captain and XO could therefore also be seen as a nice illustration of the differences between Harris's 'segregational' and 'integrational' philosophies of communication.

The Captain comes across as a mouthpiece for the 'segregational' model, which 'confuses communication with transportation; the assumption being, apparently,

[7] For Harris's views on 'distributed cognition' in relation to his own 'integrationist' perspective see Harris (2004). For discussion of the relationship between 'distributed cognition' and integrationism see Spurrett (ed.) (2004).

that information comes in neatly wrapped parcels, clearly addressed to particular destinations, and that the only communication problem is how to send them on their way and ensure their safe arrival' (Harris 1996, p. 5). Hunter, on the other hand, seems to have grasped the 'integrationist' principle that 'there is no sign without a context, and contexts are not given in advance' (Harris 1996, p. 7).

Thus, while Ramsey insists on sticking to rules that are 'not open to interpretation', Hunter opens up the closed circuit of the message transmission process and forces through a critical re-appraisal of the role and function of that process. What is meaningful at this point is not down to procedure but is now *up to us*. And it is precisely at this point that Hunter the individual appears not so much as a cog in an exquisitely distributed cognitive wheel but, respecting the mechanical metaphor, a spanner in the works.

Hunter the refusenik is also something of a *contra-disciplinarian*. Ramsey's leadership style is entirely built around a model of thought, communication and action which Love (2004) refers to as 'classical'. The chain of being passes from thought to words and from words to deeds; here the thought, there the word and its meaning, there the action. The model is indeed well suited to the military mind: the thoughts of a superior officer are encoded in words as orders, the orders are decoded unambiguously by the subordinate, and the thoughts contained therein are then enacted in behaviour. Thoughts have to do with cognition—the domain of the cognitive psychologist; the verbal encoding and decoding of thoughts belongs to the linguist; and the mundane leftovers—the performance of action—can be looked at by some third party.

Hunter takes a chainsaw to the classical model. The message does not 'contain' or 'embody' a definite thought to be derived by decoding procedures independently of action. What means what has to do with the chain of actions we are fighting to construct using all our powers of imagination about where that chain will lead us. Cognition, action and communication are not separate, 'segregated' spheres but 'integrated' aspects of a single dynamic process of activity rooted in the moral exigencies of the here-and-now. It is the fact that all our doings are simply 'forms of engagement' (Jones 2007), that is, ways of connecting us with and, thereby, affecting others in some way, that gives them their 'ineluctable moral dimension'.[8]

Thus, in opposing the Captain, the XO exposes the artificiality of Ramsey's cognitive and linguistic dogmas. His struggle is an affirmation of the fact that human thinking and action—*thinking-in-action*—is a creative activity of engagement with the big picture as we see it. That is, he asserts the right to *make meaningful signs* out of 'nothing', out of 'fragments', out of things, events and circumstances that are 'meaningless' according to 'the rules' or 'the language'. Hunter does not merely subject an already identified linguistic form or meaning to moral scrutiny. Rather, in his moral disquiet, he creatively attributes significance to something in a way that the others, at first, cannot or will not. For Hunter, what the Captain is saying is

[8]For a demonstration of the distinctive moral fabric of such mundane practical-communicative acts as carrying children versus carrying bags of groceries, see Hodges and Lindheim (2006).

defensible in terms of the artificially narrowed context of conventional procedure, certainly, but wrong when life has intervened to blow that context to pieces.

The film also teaches us that, while communicative or cognitive actions may, from a particular point of view, appear dispersed or ‘distributed’ within a network of people, machines, objects and messages etc., for an individual in the white heat of a *Crimson Tide* moment things do not feel like that at all.[9] For XO Hunter, beaten but unbowed, standing firm and stoically receiving blows to the face for his refusal to give up the vital key, he himself is what stands between the world and nuclear war. For Hunter, this is not a problem shared (and, therefore, halved) but one of his own making and entirely owned by him. Responsibility is not at all ‘distributed’ here, or ‘assigned’, but *assumed* by Hunter against all his training and professional instincts. It is his possession, his burden, his cross to bear.

Navigation with a Moral Compass

I am not claiming that Hutchins’s (1995) account has nothing at all to say about the mutiny on the *Alabama*, or that the ‘distributed cognition’ approach more generally, to which Hutchins has contributed so much, is incapable in principle of taking onboard, so to speak, the significance of such ethically charged conflicts and the role of individual agency and subjectivity within them. In any case, you may argue, it is standard Hollywood practice to exaggerate and romanticize the individual’s place in history at the expense of the culture or community to which that individual belongs and whose moral and political values that individual gives voice to. Against the individualising ideology of your average blockbuster, Hutchins is absolutely right to emphasise the cultural nature of all thinking and action and, to that end, has good reason to indulge his preference not to ‘discuss the properties of individuals before describing the culturally constituted worlds in which these properties are manifested’ (1995, p. 287).

I think he is right to emphasise that ‘seeing human cognitive activity as an integral part of such a larger system may bring us a different sense of the nature of individual cognition’ (1995, p. 287), and that ‘any attempt to explain the cognitive properties of the integral parts [i.e. individuals, PEJ] without reference to the properties of the larger system’ would be ‘incomplete’ (1995, pp. 287–288). He is also right to point out the converse: ‘Any attempt to explain the cognitive properties of such a larger system without reference to the properties of its most active integral parts would be deficient’ (1995, p. 287).

Nevertheless, it is perhaps at this very point that we begin to sense a weakness in Hutchins’s otherwise masterly treatment. The individual is nothing without the community—the ‘larger system’—and the community nothing without its individuals, true. But what do we make of this ‘larger system’? And how is the relationship between the individual and the ‘larger system’ conceived? Hutchins in fact gives us

[9]For a ‘distributed’ approach to language see Cowley (2007a, 2007b).

his definition of ‘culture’ explicitly: ‘Culture is not any collection of things, whether tangible or abstract. Rather, it is a process. It is a human cognitive process that takes place both inside and outside the minds of people. It is the process in which our everyday cultural practices are enacted’ (1995, p. 354).[10]

We may, I think, justifiably baulk at the equation of ‘culture’ with ‘cognitive process’, particularly when we learn that Hutchins’s treatment of ship navigation is an attempt ‘to apply the principal metaphor of cognitive science—cognition as computation—to the operation of this system’ (1995, p. 49) and where this computation ‘can be described in the way cognition has been traditionally described—that is, as computation realized through the creation, transformation, and propagation of representational states’ (1995, p. 49).[11]

Seen in this light, Hutchins’s account of the role of individuals with respect to this complex network of ‘representational states’ is perhaps bound to sound rather soulless: ‘I will assume that a principal role of the individuals in this setting is providing the internal structures that are required to get the external structures into coordination with one another’ (Hutchins 1995, p. 131). And Hutchins applies the same kind of thinking to the utterances made by individuals, such utterances being ‘states in structured representational media which we understand by bringing them into coordination with both external and internal structured representational media’ (1995, pp. 140–141).[12]

On reading such passages one may be forgiven for wondering if Hutchins had forgotten his own insightful critique of the ‘meaning in the message’ delusion. Just as there is no meaning, and no message either, without *somebody* (linked in many ways of course to lots of other *somebodies*) *making something meaningful* within their activity, then there are no ‘representational media’ outside and independently of *somebody’s practice* of *making something represent something else*. As Love (2004, p. 531) puts it: ‘Signhood is conferred on a sign—on what thereby becomes a sign—if and when human beings (or other semiotically competent creatures) attach a signification to it that goes beyond its intrinsic physical properties, whether in furtherance of a particular programme of activities, or to link different aspects or phases of their activities, to enrich their understanding of their local circumstances or general situation’.

[10] I’m not sure I completely understand this definition, partly because, having stated that culture is ‘a human cognitive process’, Hutchins argues in the very next sentence that ‘a major component of culture is a cognitive process’ (1995, p. 354).

[11] The reduction of ‘culture’ to ‘cognition’ which this passage implies is apparently central to the ‘distributed cognition’ approach, as this more recent formulation from Halverson (2002, p. 246) shows: ‘For me, the many phenomena of human society and activity are the result of human cognition. Much of their power arises from how cognition instantiates itself in the material world’. With such a position we have pretty much returned to the ancient, ‘idealist’ view of thought or *logos* as the source or creator of reality.

[12] Cf. Harris’s comment on papers written from a ‘distributed cognition’ perspective: ‘I note the frequency with which the catch-all term *representation* is bandied about without any serious attempt to pin it down’ (2004, p. 736).

Furthermore, as we have seen from the *Crimson Tide*, the somebody in question may decide that the 'coordination' of 'external' and 'internal' structures which is 'required' (by whom?) is, in a given situation, not on. And so I believe that it is in Hutchins's explicit formulations of the relationship between individual and cultural action that we see both the strengths and the weaknesses of his approach. The following passage, for example, makes a good, valid point about distinguishing between individual minds and the workings of the cultural system: 'If we ascribe to individual minds in isolation the properties of systems that are actually composed of individuals manipulating systems of cultural artifacts, then we have attributed to individual minds a process that they do not necessarily have, and we have failed to ask about the processes they actually must have in order to manipulate the artifacts' (Hutchins 1995, p. 173).

But at the same time the denial of the creativity of individuals' thoughts and actions implied by the reduction of their role merely to that of 'manipulating systems of cultural artifacts' is an unwelcome echo of those strains of sociological or cultural determinism (including linguistics, of course) in which human properties and powers are severed from their bearers and converted into impersonal principles of which they are the helpless subjects.

Nor, in my view, does referring to individual people as the 'most active integral parts' of the 'larger system' of culture help matters very much. If we are thinking about the 'larger system' as comprising people on the one hand and the artifacts that they create and use on the other, then, in my book, the people are not 'the most active parts', they are *the only active parts*. If, by contrast, we are thinking about the behaviour of those individuals with respect to one another, then reciprocal action and interdependence will be the order of the day although the quality of the respective contributions of individuals will not be reducible to degrees of 'activity' or 'inactivity'. It would be odd, for example, to say that John, Paul, George and Ringo were the 'most active' players in that cultural phenomenon known as 'the Beatles'.

Now, these problems with the 'distributed cognition' view have already been raised on more than one occasion. But particularly interesting in this connection is the exchange between Halverson (2002) and Nardi (2002). In her response to Halverson's exposition and defence of 'distributed cognition', Nardi puts her finger squarely on the key issue: 'Halverson notes that a serious critique of distributed cognition is that by conflating people and things, the theory denies people their humanity. The issue is raised, but I could not find a rebuttal to the critique in the paper' (2002, p. 273). A theory of human behaviour which denies what is human about behaviour—that does not sound like a good thing, does it? As Nardi goes on to say: 'the most important role of any theory about people is to declare what it means to be human' (2002, p. 274).

After these criticisms of Hutchins's rather clinical depiction of human activity and agency (or lack of it) I should say, however, that there are one or two glances towards the ethical dimension in his book. The most interesting, perhaps, is his brief description of an action that he performed himself as researcher: 'In the pilothouse I tried not to participate, but only observe. On only one occasion did I intervene, and that was a case in which I felt that by failing to speak I would put a number of people in serious danger. My intervention was a brief *sotto voce* comment to the navigator, who resolved the situation without indicating my role in it' (1995, p. 25).

If not quite at the level of the *Crimson Tide*, the episode implies the same kind of tension between professional duty and, for want of a better word, 'human' values. No doubt Hutchins could try to explain his own ethical judgement here in terms of 'coordinating' various 'representational systems' including the categories of right and wrong provided by 'the culture'. But I wonder if it felt like that at the time. And another researcher, in his place, might well have let the boat sink in the interests of ethnomethodology. . . .

Conclusion: Getting What's Coming to Us

Hutchins's work, and the 'distributed cognition' perspective generally, is designed to correct a bias in the tradition of psychology (at least in some forms of it) towards privileging individual minds (and the 'inner world') over social and cultural processes. Hutchins wants to demonstrate that what people do together they cannot do on their own and, therefore, that the actions and processes performed by the collective are not equal to the actions and processes that individuals can perform. Hutchins is right to argue that leaving out the cultural in our explanation of human thinking may lead us to believe that what is actually social is a property of the individual mind. However, it is also vital that we think carefully about what we mean by 'culture' and about sociality if we are not to convert the individual person from an active, ethically engaged creator of culture into an intersection of value-neutral 'representational states' or a 'manipulator' of morality-free cultural artifacts.

In short, there is no point in going beyond the computer metaphor—the machine metaphor—with respect to the 'inner' 'workings' of the individual mind if we simply extend the same perspective to the 'outer' 'public' workings of the team, community or culture. In other words, thinking of the culture or social interaction as a 'distributed' machine with the individual as one cog amongst many is, arguably, not much different to thinking about individual linguistic competence or thinking in terms of internal grammatical or cognitive mechanisms.

I hope, therefore, that my use of episodes from *The Crimson Tide*, this rather corny and melodramatic film, has offered a useful, and entertaining, occasion for a discussion of what often proves to be the missing link in academic treatments of language, communication, cognition and action—namely, the moral texture of our activities and dealings with one another. But how do we restore that link? How do we do justice to the socially dispersed chains of interdependence and interconnection which bind us at every step without bleaching out the agency and responsibility of individuals, without reducing people to ciphers, to mere cogs in a wheel?

On one level, this is, perhaps, about not getting too carried away or taken in by a mechanical metaphor that may have value and validity within certain contexts, or within certain limits. But, on another level, it is about the impoverished assumptions concerning the nature of humanity which particular disciplines, for particular reasons, have built into specialised, ‘segregated’, conceptual systems and methodologies.

Acknowledgements My thanks to Stephen Cowley for inviting me to the symposium and for encouraging me to write this piece. I therefore hold him morally responsible for the consequences. My thanks also to Frédéric Vallée-Tourangeau for his support and to Bert Hodges and two anonymous reviewers for their very helpful and constructive criticisms which I have tried to take onboard.

References

Cowley, S. J. (2007a). Cognitive dynamics and distributed language. *Language Sciences*, *29*, 575–583.

Cowley, S. J. (2007b). Distributed language: biomechanics, functions and the origins of talk. In C. Lyon, C. Nehaniv, & A. Cangelosi (Eds.), *The emergence and evolution of linguistic communication* (pp. 105–127). London: Springer.

Halverson, C. A. (2002). Activity theory and distributed cognition: or what does CSCW need to DO with theories? *Computer Supported Cooperative Work*, *11*, 243–267.

Harris, R. (1987). *The language machine*. Ithaca: Cornell University Press.

Harris, R. (1996). *Signs, language and communication*. London: Routledge.

Harris, R. (2004). Integrationism, language, mind and world. *Language Sciences*, *26*, 727–739.

Harris, R. (2009). *Integrationist notes and papers 2006–2008*. Gamlingay: Bright Pen.

Hodges, B. H. (2007). Good prospects: ecological and social perspectives on conforming, creating, and caring in conversation. *Language Sciences*, *29*, 584–604.

Hodges, B. H., & Lindheim, O. (2006). Carrying babies and groceries: the effect of moral and social weight on caring. *Ecological Psychology*, *18*, 93–111.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.

Jones, P. E. (2007). Why there is no such thing as “critical discourse analysis”. *Language & Communication*, *27*, 337–368.

Love, N. (2004). Cognition and the language myth. *Language Sciences*, *26*, 525–544.

Love, N. (2007). Are languages digital codes? *Language Sciences*, *29*, 690–709.

Nardi, B. A. (2002). Coda and response to Christine Halverson. *Computer Supported Cooperative Work*, *11*, 269–275.

Sacks, H. (1995). *Lectures on conversation* (Vols. I & II). Malden: Blackwell.

Silverman, D. (1998). *Harvey Sacks: social science and conversation analysis*. Cambridge: Polity Press.

Spurrett, D. (Ed.) (2004). Distributed cognition and integrational linguistics. *Language Sciences*, *26*(6), 497–501.

Stetsenko, A. (2005). Activity as object-related: resolving the dichotomy of individual and collective planes of activity. *Mind, Culture and Activity*, *12*, 70–88.

Chapter 8
Distributed Cognition at the Crime Scene

Chris Baber

Abstract The examination of a scene of crime provides both an interesting case study and analogy for consideration of Distributed Cognition. In this paper, Distribution is defined by the number of agents involved in the criminal justice process, and in terms of the relationship between a Crime Scene Examiner and the environment being searched.

The examination of a crime scene is subject to all manner of legal, ethical and scientific imperatives, and the evidence collected will be subjected to inspection by a variety of individuals with different intentions, skills and knowledge. In this paper, I will suggest that Crime Scene Examination presents an interesting and challenging domain in which to consider the notion of Distributed Cognition for the simple reason that it is not always apparent where the act of 'cognition' is situated. The ultimate aim of the criminal justice process, of course, is to acquire evidence which can be combined with information from other sources in order to produce a case that can be tried in Court. Contrary to its representation in popular fiction, the examination of a crime scene is unlikely to yield evidence that immediately links a suspect to a crime. Rather, the collection of evidence is part of a complex web of investigation that involves many individuals, each considering different forms of information in different ways. Thus, the paper begins with a cursory description of the role of the Crime Scene Examiner (CSE) within the criminal justice process.

The CSE is part of a much larger investigative system, each member of which has their own skills and roles (Smith et al. 2008). In a sense, Crime Scene Investigation involves sets of ad-hoc teams pursuing independent goals with quite limited overlap (Smith et al. 2008). Thus, there is typically a demarcation between roles. Having said this, the nature of this demarcation has been subject to significant shifting over the years, with the ongoing digitisation of Crime Scene Examination leading to further changes. For example, there used to a specific role of Crime Scene Photographer whose function was to capture and process images of the crime scene (either

C. Baber (✉)
Electronic, Electrical and Computer Engineering, The University of Birmingham, Birmingham, B15 2TT, UK
e-mail: C.baber@bham.ac.uk

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_8,

prior to evidence recovery or at stages during the recovery process, depending on the nature of the crime). However, with the growing use of digital cameras by CSEs, this role has (in some Police Forces) changed. This has the interesting implication that the function of a photograph taken by the Crime Scene Photographer was to capture the scene as clearly as possible in order to aid discussion of the scene in Court (or during subsequent investigation), but the function of a photograph taken by the CSE *could* be to illustrate the evidence recovery process; I suggest this because the capturing of images by the CSE is *part* of the activity being undertaken rather than the sole focus of the activity. Whether or not similar changes might arise in terms of specialised analysis of fingerprints, footwear marks, DNA and other evidence is a matter of continued debate. For the time being, these analyses are generally performed by Forensic scientists rather than by CSEs. This means that one of the primary roles of the CSE is the recovery of evidence and its transportation in a usable state to the laboratory of the Forensic scientist. How this recovery and transportation is performed, and how closely the Forensic scientist and CSE cooperate depends very much on the nature of the crime being examined. For much of our work, we have focused on what is called ‘Volume Crime’ (e.g., robbery, burglary), as opposed to ‘Serious Crime’ (e.g., murder, rape, kidnapping). In Volume Crime, it is likely that the recovered evidence is passed on to the Forensic Scientist via a third party (sometimes called the ‘Evidence Manager’). This means that any information pertaining to that item needs to be carefully and comprehensively recorded by the CSE prior to depositing with the Evidence Manager. It is this combined process of recovery, storing, labelling and transportation of evidence that forms the basis of several forms of computer-based CSE support (i.e., evidence management systems). Before exploring this further, we consider the archetypal detective and his approach to investigating crimes.

Sherlock Holmes and Reasoning About Crime

Sherlock Holmes tells a visiting stranger “You have come up from the South-West I see” observing that the “…clay and chalk mixture which I see upon your toes caps is quite distinctive” (Conan Doyle 1989, p. 176, *The five orange pips*). This ability to draw correct conclusions from visual evidence is one of the hallmarks of Holmes’s powers, and implies a particular form of reasoning. Holmes’s method is a form of *induction* which involves the careful observation of the environment in order to develop hypotheses and then performing a process of elimination among a number of alternative possibilities, that is, “…eliminate all other factors, and what remains must be the truth” (Conan Doyle 1989, p. 66, *The sign of four*). So that, “one simply knocks out all the central inferences and presents one’s audience with the starting-point and the conclusion, [so that] one may produce a startling, though possibly a meretricious, effect” (Conan Doyle 1989, p. 583, *The adventure of the dancing men*). He would often present his conclusions as the result of deduction (i.e., ‘Elementary, my dear Watson’) and imply that he was able to draw a conclusion from general principles to a specific observation; indeed, Holmes would often refer to his method as *deduction*. One could argue that Holmes was attempting to apply a deductive method (through his exposition of premises) but was hampered by Conan Doyle's insistence of continuing to add extra pieces of evidence, which forced him into an inductive method.

This distinction between induction and deduction is based on a broad characterization of the approaches as rival positions, namely induction as 'observations leading to theory', and deduction as 'theory guiding observation'. In reality it can be difficult to separate the two, and difficult to conceive of the 'pure' application of induction (which would involve the compiling of observations in a manner which was theoretically agnostic, and the subsequent development of a theory which was *solely* based on those observations). One would assume that observations will be, in some sense, selective and that this selectivity could be tuned by attention to specific aspects of the environment. The point of this discussion is to raise a key issue for Crime Scene Examination; there is a supposition that the work of the CSE involves the 'harvesting' of materials which would then be analysed by Forensic Scientists. CSEs are supposed to maintain neutrality in terms of collecting evidence and to conduct their work in an inductive manner, because any sense in which they are interpreting the scene could be construed as a potential for bias in the investigation. Of course, Holmes never had to face such accusations because, as a literary character, he was not guilty of bias (only of revealing the information given to him by his author) and did not have to justify his interpretations under cross-examination in Court. The question of how Crime Scene Examination treads the line between induction and deduction is explored later in this paper; before this we will consider the notions of Distributed Cognition that underlie our studies.

Distributed Cognition

The notion that cognition can be 'distributed' has been developed over the past couple of decades (Artman and Wærn 1999; Artman and Garbis 1998; Busby 2001; Flor and Hutchins 1991; Furness and Blandford 2006; Hollan et al. 2002; Hutchins 1995a, 1995b; Hutchins and Klausen 1998; Perry 2003; Scaife and Rogers 1996). While I suggest that Crime Scene Examination necessarily involves several agents performing cognitive activity, this is not to argue that this results in an 'extended mind' across these agents; as Dror and Harnad (2008) point out, to argue for an extended mind is analogous to arguing for extended migraine—just because an event occurs in one brain does not inevitably mean that other brains will share this event. Dror and Harnad's (2008) argument is that one should not separate cognitive states from mental states. This criticism raises a core problem for the notion of 'Distributed Cognition', because it implies that cognition cannot be 'distributed' across agents because one cannot share mental states. A primary assumption of 'distributed cognition' is that it is not 'cognition' which is distributed so much as objects-in-the-world, which play a role in supporting, structuring and aiding the activities of cognition.

"A main point of departure from the traditional cognitive science framework is that, at the 'work setting' level of analysis, the distributed cognition approach aims to show how *intelligent processes in human activity transcend the boundaries of the individual actor*. Hence, instead of focusing on human activity in terms of processes acting upon representations inside an individual actor's heads the method seeks to apply the same cognitive concepts, but this time, to the interactions among a number of human actors and technological devices for a given activity" (Rogers 1997, p. 2). This quotation hints at two notions of an 'extended mind'. For example, some theorists claim that the mind can become 'extended' through its interactions with the environment, for example "...certain forms of human cognizing include inextricable tangles of feedback, feed-forward and feed-around loops; loops that promiscuously criss-cross the boundaries of brain, body and world" (Clark 2008, p. xxviii). Thus, as we shall in the section entitled 'Inspection and Examination', objects-in-the-world (and the representations made of them) form resources-for-action through their ability to afford specific responses. In addition, the crime scene examination process also features a distribution of tasks. What is particularly interesting, from the point of view of Distributed Cognition, is that the process of 'find-recover-analyse-interpret-conclude' is divided between two or more people, with quite limited communication between them. The CSE might perform the 'find-recover' tasks to gather potential evidence and then submit this for the 'analyse-interpret' tasks by a Forensic Scientist, who would then pass the results on to the Officer in Charge of the case with a probability to guide the preliminary 'conclude' tasks. The Officer in Charge would then combine this evidence with other information to raise a hypothesis and add this to a Case file which would be passed to the Crown Prosecution Service. This hypothesis, if maintained, would then be tested in Court by Barristers presenting a case for and against an individual.[1] Each step of this process would be documented and conclusions drawn in such a way as to avoid potential bias.

One could draw an analogy between 'extended mind' and the debate over 'broad' and 'narrow' mental content in Philosophy. The notion of 'narrow' content might assume that a person's belief about something could be defined entirely by their intrinsic characteristics (and would not change with any changes in their environment). The notion of 'broad' content, on the other hand, is inextricably tied to the person's environment. For example, Putnam (1975) contrasted beliefs about the concept 'water' between Earth and 'Twin Earth'. Twin Earth was exactly the same as Earth, with the exception that the chemical properties of that element termed 'water' were different (although the observable properties were the same on Earth and Twin Earth). Putnam's (1975) claim was that, given identical individuals on Earth and Twin Earth, when either spoke about 'water' they would be referring to something different. This means that the intrinsic characteristics of these two identical individuals would not be sufficient to determine the meaning of the word 'water', but that there needs to be some reference to external environment. This leads Putnam

[1]This example follows the legal system in England and Wales; while other countries will follow different processes, the point is that several people are involved in the interpretation of evidence.

(1975) to make the well-known assertion that "… meanings' just ain't in the head" (p. 227).

Relating this discussion to the earlier contrast between Sherlock Holmes and contemporary CSE, we could suggest that Holmes represents the application of 'narrow' content; the world and its machinations exist solely through his (or rather, Conan Doyle's) description of them and this description cannot be challenged (simply because the stories rarely include the opportunity to develop alternative explanations). In contrast, the CSE is involved in the application of 'broad' content; the world is represented as evidence which is passed between different people who can offer different interpretations to bear on it. From this perspective, the question becomes a matter of how representations are used rather than a matter of *individual* interpretation (because these interpretations will always, in an adversarial legal system, be open to dispute).

Distributing Examination

While Sherlock Holmes provides an entertaining version of logical analysis (and serves as a template for contemporary television equivalents), his approach has many differences with modern Crime Scene and Forensic Examination. Obviously, Crime Scene Examiners do not have the benefit of the omniscient author guiding the discovery and interpretation of evidence, nor do they have the opportunity to present their findings to an informal (usually incredulous) gathering of people, as could Holmes. More importantly, Holmes's form of inductive reasoning requires the probabilistic elimination of competing hypotheses to explain a well-defined piece of evidence. The notion of a well-defined piece of evidence concerns the relationship between recognizing something as having potential evidential value and the interpretation of that evidence in terms of other information. For Holmes (and his modern, fictional counterparts) this all takes place in the head of one person; so the processes are typically assumed to involve the mental states of a single individual.

Crime Scene Examination can be considered 'distributed', in a trivial sense, in that several people are involved in the interpretation of evidence, each providing a particular perspective on this interpretation. What we see in Sherlock Holmes is a literary representation of the many-headed being of the criminal justice process in the body of a single individual. As crime scene examination grew increasingly 'scientific' so the division of tasks into discrete specialisms (each with a defined skill set) developed (Horswell 2004). Thus, it is typical for the Crime Scene Examiner and Forensic Scientist to have followed different career paths and have different skill sets (and, furthermore, for there to be a growing variety of specialisms within Forensic Science). Two further factors in the 'distribution' of Crime Scene Examination arise from the 'civilianisation' of CSE activity (the recruitment of personnel to this function from outside the Police Force) and the establishment of specific CSE units (outside the operation of separate Police stations). Each of these factors can be related to imperatives of economic and efficiency gains, but they have a bearing on how knowledge of criminal behaviour is shared and applied. For example, an understanding of criminal behaviour, gained over years of policing, could help interpret evidence; but recruiting civilian staff to these posts might remove the opportunity to gain knowledge and experience from policing. This could be dealt with through the training and exposure of new CSE personnel, or through the integration of CSE activity with other police activity. This relates to the second point, namely the removal of a CSE from local police stations to centralised services, implies the need for a means of sharing experiences and knowledge. Thus, if there is a set of similar cases in an area (say a string of burglaries with similar ways of gaining access to a building), then one would expect a link to be made between them. However, if each case is investigated by different individuals, then it might not always be possible to explore such links.

What is happening in Crime Scene Examination is the mediation of cognition through the collection, manipulation and dissemination of a variety of artifacts; each artifact is interpreted in particular ways by the agents who come into contact with it. My argument will be that, for the various agents involved in this evidence chain, each artifact can 'afford' a particular set of responses, that is, the artifacts are resources for action, and the actions will be recognized by different agents according to their training and experience. I am using the notion of 'afford' in the sense introduced by Gibson (1977, 1979), as a form of perception-action coupling in which the physical appearance of an object in the world supports particular physical responses (e.g., a pebble 'affords' grasping in the hand). Thus, the design of artefacts that are used in a work environment become changed by their use, and these changes provide cues for subsequent use (Bang and Timpka 2003; Nemeth 2003; Seagull et al. 2003). What makes this a challenging domain for discussing Distributed Cognition is that the manipulation of an artifact by one agent might have a significant bearing on the state of the artifact, which could interfere with the activity of other agents, e.g., a simple example would be the need to preserve a crime scene so as to protect evidence from contamination conflicting with the need to retrieve specific items of evidence, or the need to dust a surface to reveal fingermarks conflicting with the need to photograph the scene.

Inspection and Expectations

In their study of Crime Scene Examination, Schraagen and Leijenhorst (2001) recorded verbal protocols of the examination of a staged crime scene. They suggested, for the analysis of these protocols, that the experienced Crime Scene Examiner develops a narrative of the crime, for example considering how a person might have gained access to the building, what path they might have followed, what actions they might have performed etc. This narrative would probably be intertwined with the search activity, such that the narrative would influence the search and the search would influence the narrative. In a similar vein, Ormerod et al. (2008) suggest that "...expert investigators ... [call] ... upon internalized cognitive frames relating

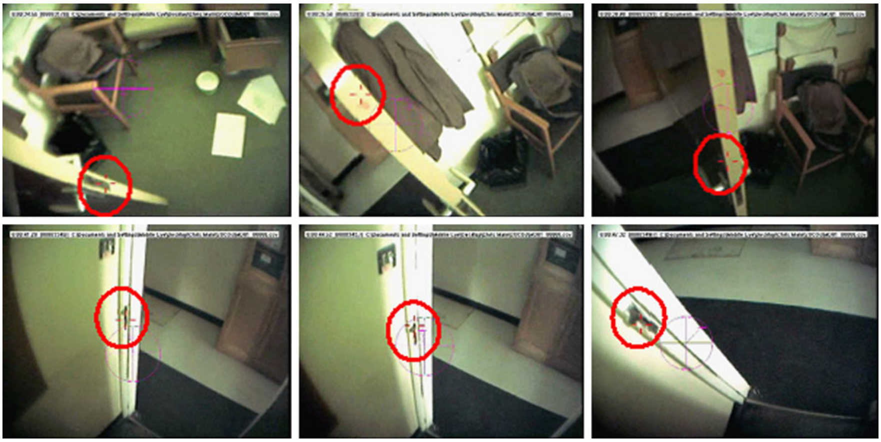


Fig. 8.1 Stills taken from mobile eye-tracker worn by Crime Scene Examiner inspecting a staged break-in (fixation indicated by cross in thick circle)

to human behaviour that allow them to generate expectations about the actions and responses of others in real time" (Ormerod et al. 2008, p. 82).

In studies using ASL MobileEye, a head-mounted eye-tracking system, we asked Crime Scene Examiners to inspect a set of staged crime scene. In one study we compared performance of three experienced Crime Scene Examiners and three Undergraduate students to search the same room under the same conditions. Of the many obvious and striking differences between the two sets of recordings, we noted that the students had a tendency to search only around locations that they believed to have links with stolen items—and so their narrative was focused solely on the loss of objects. The Crime Scene Examiners had a far more detailed narrative to guide their search and, as the stills from one recording shown below illustrate, spent a substantial part of their time looking at the door and noting possible evidence that could be recovered, for example blood stains near the latch, tool marks made by a chisel on the door frame, a footprint on the outside of the door.

Discussion with the Crime Scene Examiners showed how experience played a key role in deciding where to look for evidence and how best to examine the scene. For volume crime, the Crime Scene Examiner might walk the scene with the victim in the first instance, and then return to key locations to look for possible evidence. There was some debate as to what should be the first location to search. Standard practice might say that one begins with the Point of Entry and examines that thoroughly. In Fig. 8.1, the Point of Entry involved forcing an office door, possibly with a tool that had a sharp end, such as a chisel, which resulted in cuts around the latch. Fingermarks on the door could have been left during entry (or exit) and suggest that the entrant had cut the right thumb. Comparison between experienced CSEs and the untrained Engineering students with *no* experience of CSE work showed clear distinctions in search pattern; whereas the students all walked into the room without looking at the door, the CSEs all spent around 20 % of their total search time inspecting the door before proceeding to the rest of the room. There are two plausible explanations for this. The first is that this scene (which had been staged to replicate an office break-in) had conspicuous evidence on and around the door. However, this evidence was not so conspicuous that the students noticed it. The second is that the CSEs expect to find evidence at Point of Entry and so attend to this in detail. The CSEs, after the study, stated that this approach was 'intuitive' and 'just felt right'. Subsequent research has demonstrated that experienced CSEs focus their search on elements in the crime scene which can support recovery of useful evidence, whereas novice (student) CSEs seek to explain the crime and attend to more elements (Baber and Butler 2012). In their discussion of intuition in problem solving, Dreyfus and Dreyfus (1986) noted that "intuition is the product of deep situational involvement and recognition of similarity…; [and becomes expertise when] not only situations but also associated decisions are intuitively understood" (Dreyfus and Dreyfus 1986, p. 18). This notion is analogous to Klein's notion of Recognition-Primed Decision-making (Klein et al. 1986). In Recognition-Primed Decision-making (RPD), one can infer three broad approaches that the decision-maker might follow; (i) the situation is recognized as 'typical' and an associated set of activities would be brought to mind; (ii) the situation is defined in terms of core features, each of which would be developed in terms of (i); and (iii) the situation is unusual, and the person might mentally explore alternative strategies prior to committing to a set of activities. This study, and discussion with the Crime Scene Examiners, implies that the situation was defined in terms of (ii), and that each aspect would be considered in terms of a set of activities. The Point of Entry was explored in terms of recoverable DNA, fingermarks, and toolmarks (possibly in this order because each might be considered to have different levels of permanence and need to be recovered quickly). In a similar manner, Flin et al. (2007) have suggested that operational policing involves recognition of situations and the subsequent elicitation of appropriate response scripts, so this example of CSE suggests a three-step process by which a set of 'typical situations', such as Point of Entry, are used to guide search of a scene, which then leads to attention to items of potential evidential value, and then interpretation of these items. Thus, we could reverse Klein's RPD to describe the activity of the CSE as Decision-Primed Recognition. This is not a huge step in terms of Klein's notion of RPD because it simply follows the perception-action cycle that RPD implies: The recognition of features in the environment are responded to in terms of decisions based on previous experience, and these decision, in turn, can help shape expectations of what to look for in the environment (and to help interpret what one is looking at).

A second study concerned compared first students on a crime scene examination and forensics degree and experienced crime scene examiners. In one condition, there was a search of a ransacked office (again the scene was staged). Figure 8.2 shows a set of stills taken from an experienced Crime Scene Examiner opening the office door and immediately noticing a black mark on the floor (a), closer inspection indicates that this is a footwear mark (b) and, during the course of subsequent searching a plastic bag is found under a table and a pair of shoes found in the bag—the shoes have a black substance on their sole and the tread looks similar to that in the footwear mark (c). The scene had been staged to look as if an opportunistic

Fig. 8.2 Series of images from eye-tracking worn by experienced CSE inspecting a ransacked office

thief had broken into the office and stolen money from a petty-cash tin (which was left open on top of the desk). However, in a twist in the scenario, we had staged the scene to actually reflect an 'insurance job', that is, the office's owner had staged the crime to claim on his insurance for loss of cash, personal possessions and some computing equipment.

Most of the evidence in the scene could have been used to support the conclusion of an opportunistic crime, which was the conclusion of all 5 students and 2 of the CSEs. There were three crucial pieces of evidence which pointed to the alternative conclusion (the shoes, as shown in Fig. 8.2; the fact that the window looked to have been forced but with no obvious evidence of it being used as a point of exit, particularly as it was some 15' off the ground; the order in which the desk drawers had been opened[2]).

The stills in Fig. 8.2 show an additional aspect of the CSEs exploration of the scene. As well as being guided by their experience of likely places to search for evidence, they need to maintain a running commentary of recovered evidence so as to be able to compare subsequent finds. Interestingly, the two CSEs who did not link the shoes to the footwear mark had previously dismissed the marks as 'smudged' and 'not worth recovering'. This implies that the mark was no longer part of their running commentary and so the potential value of the shoes was not explored. The question of how a 'running commentary' is developed and indexed during a search activity could be worth further investigation. Studies of Distributed Cognition demonstrate ways in which, objects-in-the-world structure cognition. Often these objects-in-the-world are purpose-built to support specific cognitive activities, or are adapted from existing objects. Researchers would then either focus on the design of such objects, and their ability to support cognition or at ways in which activities result in the modification of objects. Crime Scene Examination represents a special case, in that the objects-in-the-world to which the person attends have been neither designed nor adapted to suit a specific cognitive activity. Rather, the objects have to be discovered by the person and then interpreted in terms of their relevance to the task of gathering evidence. In this manner, the tasks of discovering

[2] In order to prevent one drawer obscuring the contents of the next, and in order to prevent the need to close drawers, the experienced criminal is likely to open drawers from the bottom up—but in this scene, we had obviously opened them top down.

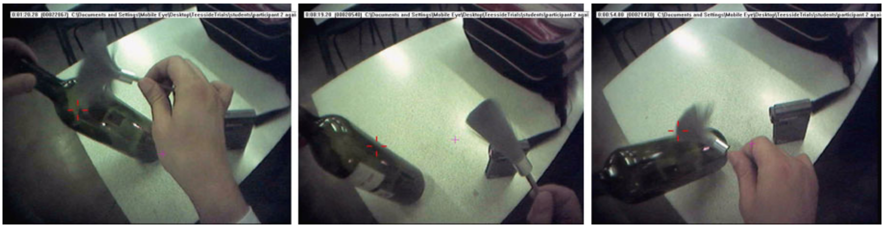


Fig. 8.3 Dusting for fingermarks

objects-in-the-world that could have evidential value can be considered a form of recognition-primed decision-making.

Evidence Recovery

As mentioned previously, one requirement of Crime Scene Examination is to select items that *could* be of evidential value. This means not only finding visible items, but also preparing surfaces so that less visible, or latent, items can be revealed. Figure 8.3, for instance, shows how a surface can be prepared to lift fingerprints. In this instance, the item being inspected (a glass bottle) is being dusted with aluminum powder using a brush. The brush is applied to the item using a swirling motion to ensure a light, even coverage. The process involved a period of brushing (for around 10 seconds), followed by a visual check (for about 5 seconds in which the bottle was gently rotated to catch light falling on any revealed marks), and then a repeated period of brushing prior to the use of tape to lift the revealed marks (or, more recently, the use of high-resolution digital photography to capture the marks) to transport them to the laboratory. In some instances, the visual check might be supplemented through the use of a handtorch which is shone orthogonally to the powdered surface. In the inspection shown in Fig. 8.3, the torch was not used but the CSE could be seen to be rotating the bottle to catch available light during the visual check phase. Concurrent verbal protocol during the search suggested that the CSE initially concentrated on two areas that were anticipated to reveal marks—and there was an assumption that each area would reveal different types of mark. Around the neck of the bottle, the search was initially for marks from fingertips and thumb holding the bottle vertically (as if carrying it) and around the middle of the bottle the search was for marks of the bottle resting across the middle of the fingers and being controlled by the thumb. Thus, a schema of how the bottle could have been used influenced the initial search.

While there are procedures in place for the recovery and analysis of finger marks, work by Dror et al. (2005) highlights how their interpretation could be biased with the provision of additional contextual information. In this study, contextual factors were manipulated by the story and photographs that were used to explain the source of the fingerprints, for example crimes with no physical harm to the person versus crimes with extreme physical harm. The study showed that in cases where the fingerprints were unambiguously different, there was little effect of context. When the fingerprints were ambiguous, namely when the certainty as to whether they were the same of different decreased, then the contextual factors seemed to play a role in increasing the likelihood of seeing a match. However, this effect was only observed for the context in which extreme physical harm featured in the background story. The study suggests that in cases where there might be some uncertainty as to whether fingerprints match and where the crime is extreme, that matching might be influenced by context. This also suggests that, while the use of a narrative to guide the collection of evidence might be beneficial, it can also bias interpretation and, by implication, search. This raises the potential (and, perhaps, often unexplored) question of how recognition-primed decisions can become biasing rather than supporting, particularly in terms of expectancy bias. This also highlights the importance of maintaining as neutral a description in crime scene reports associated with recovered evidence as possible, and shows why the inductive approach is preferable for the CSE; even if the final 'theory' to which the evidence leads is not developed by the CSE but by other people in the criminal justice process.

Evidence Sharing

The preceding discussion implies that the search of a scene is guided by experience, expectation and the ability to recognize items of evidential value. In this respect, the notion of Distributed Cognition can be interpreted in terms of the use of objects in the world as resources-for-action. The Crime Scene Examiner recognizes objects as resources-for-action which may well differ from untrained observers. For example, while the untrained observer might assume that a pane of glass in a window could yield fingermarks, they might be less inclined to immediately assume that it could also yield footwear marks, and still less inclined to recognize its potential for yielding DNA (the latter two could arise from someone climbing in through the window, or from pressing their forehead against the window to see if anyone is at home).

So far, this description looks very much like a process that involves the mental states of an individual; the CSE interprets the scene, recognizing objects as resources-for-action, and then recovers the evidence. However, what makes the Crime Scene Examination process different from a Sherlock Holmes story is that the CSE submits the evidence for interpretation by other people. Indeed, it is unlikely for the CSE's notes and reports from the scene to include any deduction. Rather the report will be as descriptive as possible. This representation, of the scene and its evidence, is passed along the recovery train. So we have a set of processes that could ostensibly represent the stimulus (or input) to a cognitive processing system. This processing is (formally) undertaken by people other than the CSE.

Once evidence has been recovered, it is placed in appropriate bags (or containers), labelled and passed on the Forensic Laboratory for further analysis. This step in the process requires some means of maintaining accurate records of who has handled the evidence, as well as the accumulation of the results of analyses. This relates to a point made earlier, that the 'distributed' nature of the Crime Scene Examination process can make this process somewhat disjointed, in that it is not uncommon for the Forensic Scientist in the laboratory to have very little information on the item recovered. One could make a strong argument that this lack of information helps an analysis to be as objective as possible, by focusing only on the item at hand (and avoiding the potential for bias that Dror et al. (2005) demonstrated). On the other hand, it might be useful to have some knowledge of the item *in situ*, so as to decide how best to conduct analysis. If the Forensic Scientist had recovered the item herself then such information would be recalled by her, but when it is delivered in a batch of bags then such information is not obviously available. As an example of why this could be problematic, consider a finger-mark left on a window. This mark might not be detailed enough to form a print, but could indicate whether the window has been forced up or whether someone climbed down the window—knowing the orientation of the mark on the window can help decide how best to analyse it, but this might not have been provided in the evidence log.

Reporting and Disclosure

In previous discussions of Crime Scene Examination, Baber et al. (2006a, 2006b) consider the manner in which narratives are passed through the evidence chain. The argument was that different people in the evidence chain develop narratives (both formal and informal) that summarise the key aspects of their interpretation of the events and environment. Thus, a victim or witness might provide an account of the events as they recall; although, of course, the nature of eye-witness testimony is notoriously contradictory and prone to error (Wells and Olson 2003). Each account would develop a particular narrative, emphasizing the aspects that the witness feels was relevant, and attempt to maintain an internal coherence and consistency (but which might differ from other accounts). Interviewing of suspects, in part, involves comparing different narratives (from the suspect versus a synthesis of the witness statements which maintains coherence and consistency). In this context, the role of forensic evidence becomes merely a tool to resolve any ambiguities in these accounts. However, of course, forensic evidence has become increasingly significant in investigations (to the extent that it is often given priority over narratives because of its assumed objectivity in comparison with the obvious subjectivity and potential for bias in the narratives). We propose that each step in the criminal justice process involves the production of narrative. There are the formal narratives that are structured by the reporting procedures and forms that are used to record investigations and analyses. This would lead to a set of reports, from Crime Scene Examiners and Forensic Scientists, which are written in a scientific style and which record details in as objective a manner as possible. Such narratives would then be subjected to scrutiny in Court in terms of the methods used to perform the analysis and the interpretation of the results. On the other hand, there are informal narratives that are passed on through discussion with agents involved in the investigation (say, between an attending officer and a victim, or between the attending officer and the crime scene examiner). These tend not to be recorded for several reasons. First, as discussed below, Laws of Disclosure mean that anything which has a bearing on the case needs to be available to both Defence and Prosecution so as to maintain fairness and balance. Second, and perhaps more importantly, much of this informal narrative could be said to involve the development of formal narrative, for example, an experienced attending officer might speak with a victim to calm or reassure them prior to taking a formal statement, and during this process the victim might have several partial accounts of what has happened but be seeking to reconcile this into a single.

The final decision of the relevance of an item of evidence is made in Court during the hearing. However, an initial assessment will be made (in the UK) by the Crown Prosecution Service which will evaluate the evidence that is being presented in support of a case and decide whether it is suitable. This raises one of the key dilemmas in evidence recovery and relates to the Laws of Disclosure. Basically, these Laws of Disclosure state that anything that has been collected as part of the investigation can be made available to both Prosecution and Defence (even if it is not presented at Court). This raises two issues for this discussion. First, the adversarial nature of the Justice System (in the UK and many other countries) means that the 'distributed cognition' involves not only cooperation and collaboration (in terms of several people contributing to a common goal) but also conflict (in terms of two parties attempting to prevent each other from achieving their goal). I am not sure that there are many other areas of distributed cognition research which come up against this problem (although, of course, one can imagine many examples from military and law enforcement). Second, the process often involves a number of different forms of analysis and interpretation. In Baber et al. (2006a, 2006b) we referred to these forms as formal and informal narratives, and suggested that there was a continual development of narratives, along several lines, over the course of an investigation and that very often these narratives might not connect.

Conclusions

In this paper, I suggest that, for Crime Scene Examination, cognition is distributed in three senses. First, there is the distribution of attention between the activities involved in searching, recovering and reporting. Second, there is the distribution of cognition between CSE personnel and the scene itself; the manner in which the scene is examined provides hints and cues to what evidence to recover, and interrupting this process (through the need to complete lengthy reports) could disrupt this process. For this activity, the environment and objects it contains, become resource-for-action that the experience and training of Crime Scene Examiners allow them to interpret in ways which might be different to that of the untrained observer. Furthermore, the manner in which recovered items are passed from one person to the next in the evidence chain can modify the role of these items as resources-for-action; each step in the process interprets the information from the previous step in terms of additional knowledge and information. Third, there is the distribution of information between CSE personnel and other people involved in the investigation. The notion of formal and informal narrative, and their development through the criminal justice process, sees these narratives as additional resources-for-action.

A 'weak' view of the Distributed Cognition argument might claim that what is being distributed is the collection of objects upon which the act of cognition can be focused. This would require objects-in-the-world to play a fairly passive role in the process of cognition and for them to function as vehicles for the storage or representation of information. The artefacts allow users to off-load information (Scaife and Rogers 1996) and also a record of previous activity. In this version, the objects have their states altered by the actions that their users perform on them (e.g., through note-taking, folding or other markings). Furthermore, not only do these objects provide a means of recording and storing information, but their design affords (or influences) the actions of the person using them.

A 'strong' view of Distributed Cognition, posits that it is the tasks involved in cognition which are being distributed. One way in which the activity of the CSE differs from some of these domains, is in the initial definition of objects-in-the-world, and for these objects to be 'revealed' in order to be recovered. This would regard the role of the CSE is primarily one of induction, or rather, as one of providing the set of alternatives upon which a process of induction could be applied. I would suggest that the act of induction takes place in the Court (or at least in the Crown Prosecution Service which decides whether a Case can be presented to Court). Prior to this act of induction, there are initial acts of deduction which are formally assigned to the Forensic Scientists, in their analysis and interpretation of evidence, but also informally applied by the CSE in the decision as to where to look and what to recover. In this view, one would expect agents and objects-in-the-world to be more active and capable of either performing, or at least participating in, information processing tasks. For example, Hutchins (1995b) famously speaks about the ways in which the flight-crew and their instruments work together to monitor the speed at which an aircraft is flying; his assertion is that this knowledge does not reside in the head of one specific individual, but is derived from the collection of information that is available in the cockpit. Perhaps, a point to note here is that, ultimately, there needs to be some 'cognizing entity' that is capable of combining the various bits of data into a coherent 'whole' and that this requires a set of mental capabilities that are uniquely human.

Both views raise questions that relate to the manner in which cognition becomes a matter of sharing tasks. In terms of distributed cognition, the work reported in this paper covers both the 'weak' and 'strong' views of distributed cognition. From the 'weak' view, it is argued that the training, knowledge and experience of Crime Scene Examiners allow them to use the environment and the artefacts within it, together with the collection of narratives through the criminal justice process, as resources-for-action in a manner that might be alien to the non-expert. In this way, the Crime Scene Examiner will not only search for specific artefacts but also be able to identify locations which could yield non-visible materials (e.g., places to check for fingerprints, DNA and other evidence). The use of eye-tracking and verbal protocol from crime scene examination show how the approach to searching a scene differs with experience. This has also formed the impetus for design and development of wearable computers to support CSE (Baber et al. 2009; Cross et al. 2007). From the 'strong' view, the reporting and interpretation of evidence from a crime scene through the criminal justice process implies a collective activity (which might not be coordinated by a central agency) that accumulates information to a point at which its interpretation can be tested in Court. While neither approach should be taken to imply that mental states are distributed across individuals, both imply that the action of one individual will form the basis for actions of the next. In this manner, the criminal justice process is able to 'know' the collected evidence, even though it is unlikely that a single individual will have access to all of the information collected during the examination.

Acknowledgements The work reported in this paper was supported by a grant from EPSRC [GR/S85115 MsSAM (Methods to Support Shared Analysis for Mobile Investigators)]. I am grateful to all of the people who worked with me on this project, and to the Crime Scene Examiners who gave an insight into their work.

References

Artman, H., & Garbis, C. (1998). Situation awareness as distributed cognition. In T. R. G. Green, L. Bannon, C. P. Warren, & J. Buckley (Eds.), *ECCE 9: proceedings of the ninth European conference on cognitive ergonomics* (pp. 151–156). Le Chesnay: European Association of Cognitive, Ergonomics (EACE).

Artman, H., & Wærn, Y. (1999). Distributed cognition in an emergency co-ordination center. *Cognition Technology & Work*, *1*, 237–246.

Baber, C., & Butler, M. (2012). Expertise in crime scene examination comparing search strategies of expert and novice crime scene examiners in simulated crime scenes. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*, 413–424.

Baber, C., Smith, P., Cross, J., Hunter, J., & McMaster, R. (2006a). Crime scene investigation as distributed cognition. *Pragmatics & Cognition*, *14*, 357–385.

Baber, C., Smith, P., Panesar, S., Yang, F., & Cross, J. (2006b). Supporting crime scene investigation. In N. Bryan-Kinns, A. Blandford, P. Curzon, & L. Nigay (Eds.), *People and computers XX—engage* (pp. 103–116). London: Springer.

Baber, C., Cross, J., Smith, P., & Hunter, J. (2009). Mobile technology for crime scene investigation. *International Journal of Human-Computer Studies*, *67*, 464–474.

Bang, M., & Timpka, T. (2003). Cognitive tools in medical teamwork: the spatial arrangement of patient records. *Methods of Information in Medicine*, *42*, 331–336.

Busby, J. S. (2001). Practices in design concept selection as distributed cognition. *Cognition Technology & Work*, *3*, 140–149.

Clark, A. (2008). *Supersizing the mind: embodiment, action and cognitive extension*. Oxford: Oxford University Press.

Cross, J., Baber, C., & Smith, P. (2007). Multi-platform crime scene investigation tool. In *The 11th international symposium on wearable computers* (pp. 55–62). Santa Monica: IEEE Comput. Soc.

Doyle, A. C. (1989). *The original illustrated Strand Sherlock Holmes*. Ware: Wordsworth Editions.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine*. New York: Free Press.

Dror, I. E., & Harnad, S. (2008). Offloading cognition onto cognitive technology. In I. E. Dror & S. Harnad (Eds.), *Cognition distributed: how cognitive technology extends our minds* (pp. 1–23). London: Benjamins.

Dror, I. E., Péron, A., Hind, S.-L., & Charlton, D. (2005). When emotions get the better of us: the effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology*, *19*, 799–809.

Flin, R., Pender, Z., Wujec, L., Grant, V., & Stewart, E. (2007). Police officers' assessment of operational situations. *Policing: An International Journal of Police Strategies & Management*, *30*, 310–323.

Flor, N. V., & Hutchins, E. L. (1991). Analyzing distributed cognition in software teams: a case study of team programming during perfective software maintenance. In J. Koenemann-Belliveau, T. G. Moher, & S. P. Robertson (Eds.), *Empirical studies of programmers: fourth workshop* (pp. 36–64). Norwood: Ablex Publishing Corporation.

Furness, D., & Blandford, A. (2006). Understanding emergency medical dispatch in terms of distributed cognition: a case study. *Ergonomics*, *49*, 1174–1203.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing*. Hillsdale: Erlbaum.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Hillsdale: Erlbaum.

Hollan, J., Hutchins, E., & Kirsch, D. (2002). Distributed cognition: toward a new foundation for human-computer interaction. In J. Carroll (Ed.), *Human-computer interaction in the new millennium* (pp. 75–94). New York: Addison-Wesley.

Horswell, J. (2004). *The practice of crime scene investigation*. London: CRC Press.

Hutchins, E. (1995a). *Cognition in the wild*. Cambridge: MIT Press.

Hutchins, E. (1995b). How a cockpit remembers its speed. *Cognitive Science*, *19*, 265–288.

Hutchins, E., & Klausen, T. (1998). Distributed cognition in an airline cockpit. In Y. Engeström & D. Middleton (Eds.), *Cognition and communication at work* (pp. 15–34). Cambridge: Cambridge University Press.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. In: *Human Factors and Ergonomics Society 30th annual meeting* (pp. 576–580).

Nemeth, C. (2003). How cognitive artefacts support distributed cognition in acute care. In *Proceedings of the 47th annual meeting of the Human Factors and Ergonomics Society* (pp. 381–385). Santa Monica: Human Factors and Ergonomics Society.

Ormerod, T. C., Barrett, E. C., & Taylor, P. J. (2008). Investigating sensemaking in criminal contexts. In J. M. Schraagen, L. G. Militello, T. Ormerod, & R. Lipshitz (Eds.), *Naturalistic decision making and macro-cognition* (pp. 81–102). Avebury: Ashgate

Perry, M. (2003). Distributed cognition. In J. Carroll (Ed.), *HCI models, theories and frameworks: towards and interdisciplinary science* (pp. 193–223). San Francisco: Morgan Kaufmann.

Putnam, H. (1975). *The meaning of meaning. Language, mind and knowledge: philosophical papers* (Vol. 2). Cambridge: Cambridge University Press.

Rogers, Y. (1997). A brief introduction to distributed cognition. http://mcs.open.ac.uk/yr258/papers/dcog/dcog-brief-intro.pdf.

Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work. *International Journal of Human-Computer Studies*, *45*, 185–213.

Schraagen, J. M., & Leijenhorst, H. (2001). Searching for evidence: knowledge and search strategies used by forensic scientists. In E. Salas & G. Klein (Eds.), *Linking expertise and naturalistic decision making* (pp. 263–274). Mahwah: LEA.

Seagull, F. J., Plasters, C., Xiao, Y., & Mackenzie, C. F. (2003). Collaborative management of complex coordination systems: operating room schedule coordination. In *Proceedings of the 47th annual meeting of the Human Factors and Ergonomics Society* (pp. 1521–1525). Santa Monica: Human Factors and Ergonomics Society.

Smith, P. A., Baber, C., Hunter, J., & Butler, M. (2008). Measuring team skills in crime scene examination: exploring ad hoc teams. *Ergonomics*, *51*, 1463–1488.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*, 277–295.

Chapter 9
Socially Distributed Cognition in Loosely Coupled Systems

Mark Perry

Abstract Distributed cognition provides a theoretical framework for the analysis of data from socio-technical systems within a problem-solving framework. While the approach has been applied in tightly-constrained activity domains, composed of well-structured problems and highly organised infrastructures, little is known about its use in other forms of activity systems. In this paper, we explore how distributed cognition could be applied in less well constrained settings, with ill-structured problems and loosely organised resources sets, critically reflecting on this using data from a field study. The findings suggest that the use of distributed cognition in an augmented form can be useful in the analysis of a wide range of activity systems in loosely coupled settings.

The question of how thought is distributed over a variety of non-neurological resources has received a considerable amount of interest, both within the communities of cognitive science and beyond to anthropology, sociology and even computer science. Yet while theoretical frameworks are being developed and tested to identify the socio-cognitive mechanisms through which this externalised form of cognition is organised and co-ordinated, this body of empirical research is typically limited to relatively narrowly bounded systems of information representations and agents. At the same time, the sorts of problems that it has examined are tightly defined (i.e., there are clearly articulated and specific goals), so that the problem solving agents are aware of what the state of the final problem resolution will be. This is important and groundbreaking work for understanding human behaviour and thinking 'in action'. However, many problem-solving situations are not so well structured and resource-delimited. This includes settings in which problem solving is distributed over technical, social and organisational systems that do not have well defined boundaries, so that multiple computational structures may be employed, or where the problem solving resources available to these systems may dynamically change: agents can co-opt a range of tools into their work and new agents may be

M. Perry (✉)
Department of Information Systems and Computing, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK
e-mail: Mark.Perry@brunel.ac.uk

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*, DOI 10.1007/978-1-4471-5125-8_9,

introduced into the system. Similarly, there are problem-solving situations in which the problem is itself not clearly defined or underspecified (these are known as ill defined or ill structured problems, Simon 1973), and part of the problem solving activity is to determine the form of the solution. We call these 'loosely coupled' systems.

The importance of loosely coupled systems is that are they are typical of many kinds of problem solving behaviour seen in socially organised collective activity and organisational life. Understanding problem solving within these system is therefore a matter of great practical concern for their designers and managers. However, and as we will demonstrate in the paper, the analytic techniques that have been developed to understand problem solving in tightly delineated systems undertaking well specified activities are not easily tuned to these loosely coupled systems. Deriving an analysis of such activity directly from the methods used in existing studies may therefore prove to be problematic.

One of the most clearly articulated approaches used to examine distributed problem solving that extends cognition into artefacts and social relationships is known as distributed cognition (DCog; Hutchins 1995a, 1995b). With some variations in its application (e.g., Wright et al. 2000; Blanford and Furniss 2006), it is perhaps the most clearly articulated, critiqued, commonly used and well known form of exploring how distributed action can be examined as a cognitive process (see, for example, Giere 2004), and it is for this reason that we have chosen to re-examine its use here. Distributed cognition draws its theoretical and analytical foundations from cognitive science, augmenting these with techniques for primary data collection and ethnographic representation drawn from sociology and anthropology. It is unusual in that it seeks to study people, social interaction and artefact use within a common framework. The focus of DCog is on the representational transformations that occur in external media. In practice, information about the settings is usually captured using ethnographically informed observations and interview-based methods. This has been described as 'cognitive ethnography' (Hutchins 1995a). Here, attention is concentrated on the agents (who co-ordinate the representational transformations) and the artefacts in the environment (media that encode the physical representations of the system's internal state). The interaction between these elements is crucial in understanding how representational transformations are co-ordinated. We distinguish this 'flavour' of distributed cognition developed by Hutchins and drawing from cognitive science as DCog, and differentiate it from other uses of the term. As noted by Halverson (2002), there are problematic issues in the various interpretations of the term 'distributed cognition', which is used in an analytically different way by writers such as Heath and Luff (1991), and in the book 'Distributed cognitions' (Salomon 1993), or simply as a synonym for socially shared cognition (see Engestrom and Middleton 1996; Resnick et al. 1991).

While DCog has been successfully applied in understanding some working environments, its application has been limited to a tightly constrained set of problems, with a pre-specified set of active agents and a limited set of representation bearing artefacts. As McGarry (2005, p. 187) notes, these analyses have taken place in 'highly rationalised environments' precisely because the idea of 'social systems as computation' maps most closely onto these types of setting as a solidified cognitive architecture with an easily identifiable end-point. As it stands, DCog has proved a useful and insightful approach to examining these settings, and the focus of the approach (as applied in these studies) matches well with the structured settings of this work. Yet, as we have noted, characteristics such as these are not shared in a vast range of work settings, such as team-based game playing, design, financial services, or medical work. It is interesting to note here that the existing DCog analyses have focused on partial elements of wider work settings (e.g., in navigation, Hutchins 1995a; air-traffic control, Hutchins 1995b; Halverson 1995; fMRI brain mapping, Alač and Hutchins 2004), emergency control centres (Artman and Wærn 1999) and telephone call centres (Halverson 2002). The nearest that DCog analyses have come to this in their application to highly dynamic and loosely structured settings are Hazelhurst's (1994) and Holder's (1999) PhD theses, although these do not critically examine the methodological extensions that have been made to extend the approach. Moreover, the problems described within even these settings have been carefully 'bounded' to create a simplified problem space that does not account for the larger setting.

As an example of this, we put forward ship navigation as the archetypal complex and socially supported problem that DCog analyses have sought to examine. Drawing from Hutchins's (1995a) study of navigation on a US naval vessel this is, in his own words, 'a partially closed information loop' (p. 39):

> "The activities of the Navigation Department revolve around a computational ritual known as the *fix cycle*. The fix cycle has two major parts: determining the present position of the ship and projecting its future position. The fix cycle gathers various bits of information about the ship's location in the world and brings them together in a representation of the ship's position." (1995a, p. 26)

This process involves taking visual bearings that can be translated into lines that should intersect on a navigational chart. At the risk of simplifying such a complex operation, this begins when two 'pelorus operators' use a telescope-like device (alidades) to take visual bearings of landmarks, and assign a precise directionality to these bearings ('lines of position'). This information is communicated to the bearing recorder and plotter in the chartroom who mark this on the chart as a pencil line using a 'hoey' (a one-armed protractor) to map the lines of position accurately onto the chart. All of these tasks need to be done on a highly fixed timeline, in the correct order, and parts of this can only be undertaken by particular people. Various technologies are employed in the fix cycle (e.g., charts, pelorus, hoey) and it involves six team members performing various specialised tasks. Where breakdowns occur in the standardized navigational process, 'the team are able to compensate for local breakdowns by going beyond the normative procedures to make sure that representational states propagate when and where they should' (Hutchins 1995a, p. 228). Nevertheless, despite micro-scale adjustments in the tools used and organisational arrangement between the system's components, the end result of the fix cycle is the same (a set of lines of position on a chart), and the group is unable (or at least in the analyses presented) to enrol additional resources outside those that already form the existing set.

Limiting this scope, of course, has benefits in providing a focus to the analysis. By artificially setting the problem situation up as a 'closed' system, the DCog analyst does not need to face the problems of extending his or her analysis outside of these limited settings. While the analyses that existing studies have developed are extremely valuable, extending the domain of interest beyond the limited systems explored would provide an interesting and valuable insight into the ways that these sub-systems were embedded into the larger work systems of the workplace. So, developing Hutchins' (1995a) work on navigation for example, we might learn how navigational information is interpreted in the other functions of the ship, and how other aspects of activity on the ship might impact back into the navigational subsystem. This type of analysis poses a different set of problems to those examined in these previous studies, but such analyses would nevertheless be interesting and useful to understand wider systems of context activity than the simple(r) activity in focus. Indeed, Hutchins does touch on this as a practical concern, stating that 'low-level functional systems may be embedded in larger functional systems' (p. 153), although in the cases described, even these larger functional systems as described are fairly restricted in their complexity and reach.

In the following sections, the foundations and existing use of DCog will be discussed and examined to highlight the major concerns over its use in large and complex organisational settings. Briefly reviewing the theoretical and methodological underpinnings of the DCog framework will allow us to identify and discuss the areas of difference between tightly and loosely coupled systems. To demonstrate this, we then draw on a field study, pulling out findings to highlight the application of DCog and to motivate suggestions for the development of DCog and its evolution as an analytical method.

Distributed and Socially Distributed Cognition

DCog in Cognitive Science The theoretical framing and terminology used in DCog reflect its roots in cognitive science. Classical cognitive science provides a conceptual framework with which to examine intelligence and problem solving, exploring how information is represented in the cognitive system, and how these representations are transformed, combined and propagated through the system in goal-oriented behaviour (Simon 1981). These cognitive processes and representations are normally considered to be internal to the individual and held mentally. In DCog, a larger granularity of analysis is selected than that of the individual, to include the use of information artefacts in the cognitive process (see, *inter alia*, Hutchins and Klausen 1991; Norman 1993; Hutchins 1995b). These systems include artefacts such as a pen and paper, which can be used as memory aids, as well as more complex artefacts that are integral to the computation, such as slide rules. As such, DCog provides a coherent framework with which to structure the analysis of activity systems in terms of their informational content and problem solving activities. As well as incorporating artefacts in the cognitive analysis, DCog also differs from classical cognitive science in its views of the cognitive process. DCog suggests that the cognitive process can be described as being distributed over a number of people co-operating through social mechanisms, often referred to as 'socially distributed cognition' (Hutchins 1995a; Hollan et al. 2001).

The unit of analysis in DCog may consist of any number of representations, embodied in people, computerised artefacts and non-technological artefacts (Rogers and Ellis 1994). As a framework, DCog provides a unique insight into how technology and the socially generated media of communication act upon and transform representations, and in doing so, perform computations, or information processing activity. The aim of DCog is therefore to understand how intelligence is manifested *at the systems level* and not the individual cognitive level (Hutchins 1995a). The mix of people and artefacts in a given situation contribute to the *functional system* of activity, which includes all of the representation-carrying and representation-transforming entities involved in the problem solving activity. A distributed cognitive analysis examines the means by which the functional system is organised to perform problem solving. From a computational perspective, the functional system takes in inputs (representations), and transforms these representations by propagating them around the units of the system. A distributed cognitive analysis involves deriving the *external* symbol system (cf. Newell and Simon 1972) by capturing the elements of processing (representations and processes) that transform system inputs into outputs for particular tasks. In many cases, these distributed representations and processes are brought together by agents—people—and co-ordinated through social mechanisms.

The Division of Labour Socially distributed cognition allows the analyst to describe group activity as a computation realised through the creation, transformation and propagation of representational states (Hutchins 1995a; cf. Simon 1981). By bringing representations in the system into co-ordination with each other, information can be propagated through the larger cognitive system, being continually modified and processed by a number of individuals and artefacts, until a desired result is reached. However, while processing of the information available to the group is analogous to an individual's cognitive capabilities, the architecture of this activity differs significantly. How these socially embodied activities are organised with respect to one another is known as the 'division of labour'. Understanding how this division of labour operates within the functional system provides a means of understanding its internal organisation and working practices—in effect it determines a large component of the *computational architecture* of the distributed cognitive system. Of course, there are many ways that work *can* be done, and this division of labour is malleable and subject to change as the agents within the functional system learn, change roles or introduce new technologies into their work.

Distributing work across a group of agents must involve some form of organisation to co-ordinate activity to develop a working division of labour, or an architecture for computationally acting on the problem representation(s) to affect a solution. To solve a problem collaboratively, this division of labour must operate so that work is broken into parts but managed so that these (partially resolved) components can be re-incorporated. Internal cognitive (i.e., mental) factors that are effectively invisible to the group (and to the analyst) need to be trained on the problem by individuals to bring their expertise to bear on these sub-tasks before they are able to re-incorporate their sub-task with the global task. This means that within the distributed cognitive system, problem-solving expertise lies not only in the knowledge and skills within individuals, but in the organisation of those individuals, and on their ability to co-ordinate their work with one another. This larger unit of knowledge organisation may be determined explicitly in predetermined protocols and procedures, or more informally through the context of their work environment being made visible or through the configuration of the artefacts used in the environment. DCog thus presents a method of describing and analysing how this overall system operates, using a computational metaphor applied to the functional system as a whole.

The extent to which the division of labour is clear-cut and explicit (i.e., visible to the analyst) in the work activity will determine the extent to which analysts can develop a description of the distributed cognitive system. Where the activities, agents and information artefacts in a workplace system are constrained and work roles are clearly specified in known protocols, the application of DCog to analyse the functional system should be relatively straightforward. This is likely to be particularly the case if the problem representation is explicit and can be tracked through its transformations in the system. Unfortunately, as any experienced ethnographer will recount, this is rarely the case, even if work with systems appears to be well defined at first impression. In many cases, the only way to find systems that are even remotely well defined is to pick out (relatively) simple sub-activities from workplaces and to examine these—as has been the case for the majority of DCog studies to date.

DCog: Application in Data Collection and Analysis As an analytic framework, DCog focuses the analysis on the salient points relating to the cognitive (i.e., information processing) characteristics of the functional system to structure the data available. Thus, data collection within the analytic framework of DCog must allow the identification of the representations and processes in the goal-directed behaviour. The analyst needs to show how these elements are used as resources in information processing by demonstrating how this is mediated through *action*. Emphasis is placed on the role of the artefacts carrying information representations and on collaboration around these artefacts. Analysis therefore centres on the artefacts that are used or created, how they are used, who they are used by, how changes are made to them, and how the organisation structures access to these artefacts. Amongst other tasks, this may involve mapping out the information flows through the organisational structure, identifying the sources and sinks of this information, the tools used to manipulate and transmit it, and the 'chains of command' initiating activities.

While the framework of DCog is non-prescriptive in its application to data collection, one method—ethnography—has come to be pre-eminent. Ethnography has its own traditions within sociology and anthropology and has long been used as a technique for gathering naturalistic data about activity within workplace and other settings. However, at its core, ethnographic analysis provides a means of exploring *how* human activity is organised (Hughes et al. 1992; see also Hammersley and Atkinson 1995 for an overview). It allows the analyst to physically 'enter' the functional system and to build up a picture of activity from the member perspectives, but which encompasses a higher-level perspective than the individual, showing interdependencies within the division of labour. These may be invisible to the participants themselves.

In principle, the analytical framework offered by DCog is clearly of value in its analysis of loosely coupled systems, through demonstrating the mechanisms of co-ordination and collaboration in problem solving. Its focus on the 'social' aspects of information processing identifies it clearly as different from almost all other forms of analysis as it uses the same 'language' to describe all of the system components and their operation.

Comparative Analysis: Examples from a Loosely Coupled System

This section introduces a field study from which several examples are drawn, and provides a basis from which we argue for the extension of DCog as an analytic approach to the study of loosely coupled systems. Its intention is to demonstrate (i) the similarities and differences between the traditional domain of study for DCog and less constrained, loosely coupled organisational settings; and (ii) an approach to gathering data that allows us to undertake an analysis of aspects of activity within these settings.

Background to the Study Setting The fieldwork was conducted as part of a larger project examining the use of information technology in construction, and was conducted over the course of around a year by the author who had no previous knowledge of construction. A number of field visits of two-week duration were taken over this time, as the author was embedded into the teams. This was largely non-intrusive, as team members were used to apprentices and other staff shadowing them. The field study involved an examination of design and construction on a large road-building project (see also Perry 2003) that took place over a three-year period. Fieldwork covered the participation of five (fluctuating in number) spatially distributed teams employed by a construction company. To build the permanent road structures, the construction company had to design non-permanent, supporting structures known as 'temporary works' (concrete moulds, supply roads, and so on). Temporary works are derived from the designs of the permanent works, but their designers have to take into account the site conditions (slope, weather, soil condition, and existing structures) and the available resources (money, time, materials, equipment and labour) that are not documented in the permanent works designs.

Several people were involved in the functional system developing the temporary works design, including a construction team, made up of a team leader, various engineers, foremen and gangers (work supervisors), and general labour. The team operated on the site, but were supported off-site in specifying the design of the temporary works by a design co-ordinator who worked at another location on the building site. The design co-ordinator worked closely with a design engineer (located several miles away from the site) to draw up a temporary works design. A number of other individuals and stakeholder groups external to the organisation were also closely involved in the process, including a ‘resident engineer’ (RE), who checked that the road was designed and built to contractual standards.

The temporary works design processes were partially prescribed in the handbooks and manuals that determined relationships between people, and in the organisational structures they inhabited. These specified where responsibilities for tasks lay and determined the roles that the individuals had to fulfil. While these were generally followed, particularly the safety-related rules, they were used more as organising principles, followed when appropriate, but worked around when other methods might be more appropriate in the circumstances. Structures in the environment at the site and office also played a role in determining the processes that would be applied to particular problems. These determined the configurations of representations and processes that *could* be applied to the design problem. For example, the physical size of the site made locating people difficult, and communications were complicated accordingly. However, when in the construction team’s on-site office, the engineers and foremen shared a physical space and had a wide range of media with which to communicate.

In the terminology of DCog, the transformational activity in temporary works design involved taking inputs into the functional system, transforming them into representations and re-representing them until they match the desired goals. However, the different resources available in particular circumstances, and the different ways in which individuals could resolve the problem-solving situation meant that problem resolution could be achieved in a variety of ways. Hutchins (1995a) also describes this, showing how ‘local functional systems’ are built up around—and exploit—the particularities of individual problem situations.

Illustrative Examples from Fieldwork

Having established the broad stetting of the field study, we can now look at it in relation to parts i. and ii. above, exploring the differences between traditional studies of DCog and the construction setting, as well as the approaches used to gather data in undertaking an analysis of aspects of construction work within the DCog framework.

Access to Resources The key resources in design and construction are rarely predetermined before the problem solving activity is initiated. The fieldwork illustrates how the problem solving activity changed as additional agents and artefacts become available. When new designs or requests for information arrived at the site, there were a variety of different follow-on activities that might result from this, although it was by no means ‘programmatic’ what these might be. Hutchins (1995a) recognises this ‘availability of data’ as a key controlling factor in how a functional system

Fig. 9.1 The construction team office

can organise its computational and social structure: additional information can lead to new social arrangements, which in turn lead to new computational structures (Hutchins 1995a, p. 342). Where Hutchins's analysis diverges from the concerns relevant to our fieldwork is that these new computational structures arise from the time and frequency of fix cycle data that is available and the agents available for acting on it, while in construction, new computational structures are far more substantive and may include the arrival of a different form of (relevant) information and/or new agents with entirely new roles and responsibilities into the system. This is a fundamental difference with wide reaching consequences: it is not simply the local configuration of the functional system that can change over time, but its very constitution.

This can be illustrated with our fieldwork. The construction team's office was an important resource in their work. It had an 'open plan' layout, and the engineers and quantity surveyors were able to see who else was present. It allowed them to speak to each other without moving from their desks, to overhear other people on the telephone or when speaking to each other, and to see the information laid out on each other's desks. While the construction team were centred in this office, individuals spent a large amount of their time on-site, and the distributed nature of the construction site made contacting these individuals difficult. When people were not present to talk to directly, other media were used to communicate, either synchronously through the use of the radio link, or asynchronously, through placing written notes, sketches, method statements or risk assessments on people's desks, or jotting notes onto a whiteboard. Messages were also left with people who were in the office for when that person came back. A photograph of the office space, its contents and layout is shown in Fig. 9.1.

As can be observed, the workplace was informationally very rich. Paper covered almost every surface, often several layers deep, even pinned onto the walls. When information was required from a person who was not physically present, this 'desk litter' could provide clues to their location, in the forms of the drawings and other representations on the desk, indicating the task that they were currently engaged in, and providing a guide to their current location. Other artefacts also provided information about the whereabouts of people: if a person's Wellington boots and hard hat were missing, they were probably out on site; if someone had a pair of muddy boots under their desk, it meant that they had been on the site and could be asked about the current work situation. Depending on the weather, it was even possible to see how long ago a person had been out on the site, for example from the wet or dried mud on boots, which could be useful if one of the team was trying to locate another individual out on the site. Equipment such as the geodometer was also useful in this way—if it was missing from the office, then a graduate engineer would probably be out on site with it and could be asked to run a favour over the radio. Even the window was used to see whether a person's vehicle was in the car park outside the office: if this was the case, then that person was highly likely to be somewhere on the site.

Spoken communication was conducted from the desks, allowing all of the participants in the room to be aware of developments, or allowing them to contribute to the discussion. When the senior or site engineers wanted to speak to the graduate engineers, they would stand up and chat over the tops of the partitions, providing a visual and auditory focus of attention in the room. This allowed people to work while keeping an ear to the conversation and keeping abreast of developments, to ask questions, but also allowing them to enter the conversation and add to the discussion. Within the literature of workplace studies, such observations are relatively commonplace (e.g., Heath and Luff 1991; Murray 1993). However, within the DCog framework, the analyst must draw from his or her observations how particular representations are propagated and transformed. Here, in the example above, the potential range of representations that participants could select from was vast, and they could be used in a variety of ways to achieve the same goal. In the event of a similar situation arising, it was unlikely that the same resources would be available in the same configurations as before, and used in the same way. This differs significantly from the examples of DCog documented in aircraft and navigational settings, where stable behavioural patterns (Hutchins calls these 'computational sequences') tend to be recurrent.

A consequence of this resource flexibility is that it is harder to build generalisations or models about problem-solving activity. In these situations, ethnographic descriptions of activity are likely to represent 'one-off' solutions, generated by members who generate and maintain a computational structure that utilises only a subset of the potential resources available. Such descriptions will therefore be useful to illustrate how and which resources are used, but they cannot be seen as anything more than exemplars, or instances of problem solving. This is informative in understanding the activity *in general*, but not so much the predictive performance of *an* activity. So, for example, it may be useful in understanding the formalisms within engineering drawings that allow people to communicate using them, but not in how a particular problem might be resolved in which engineering drawings were a component part.

Problem Structure The problems faced in construction were often poorly understood prior to the initiation of the problem solving activity (i.e., they were ill-structured), and a component of this activity was to understand exactly what the problem was as well as how to resolve it. For example, much of the physical component of construction work was demand-led, and work could only occur when the site had been prepared: materials or other resources might have to be ordered or cancelled at the last minute because the site was prepared for the work earlier or later than expected. The use of different construction methods or materials (arising from product non-availability or particularities of the situation) in the permanent structures could thus change the project's specifications. This differs from previous studies of DCog in navigation and cockpit activity, in which relatively regularised and well-structured problems are encountered: thus there is the fix cycle in navigation and the landing sequence in an aircraft cockpit. Deviations can and do occur, but these are relatively highly constrained and delineated within a well understood set of criteria—the goal is clear, and the set of operations that the participants have to operate on is limited and practised (although see Hutchins 1991). The examples given below demonstrate the structure of the problem faced by the designers at the construction site. The section begins with a description of the global design situation and shows how that within this, ongoing problems were identified and transformed from ill-structured problems to well-structured problems.

In official project procedures, the team's senior engineer should formally present the initial 'design brief' (a temporary works design specification) to the design co-ordinator before a temporary works design could be generated. In practice, this design brief was often little more than a few ideas sketched or jotted onto a scrap sheet of paper. This occurred because the senior engineer had too little time to perform the task, and often had very little understanding of what information the design engineer might require in the problem specification. Through discussions with the design co-ordinator, a detailed specification would be generated, containing information about the site conditions, the materials, labour and other resources available to construct the temporary works structure. The construction team's senior engineer and the design co-ordinator would then pore over the permanent works drawings, the initial temporary works design brief and several sheets of blank paper. Here's an example:

> *Senior engineer* (*SE*): '*If you look here, there's a barrel run there*' (*points at sketch generated in the meeting of a section view through a design structure*)
>
> *Design co-ordinator*: '*Yes I see*'.
>
> *SE*: '*So if we dig here…*' (*he holds one hand to the sketch and runs a finger on the other hand along a permanent works drawing* (*plan view*) *beside the sketch, indicating a line of reference*)
>
> *Design co-ordinator*: '*No you can't do that because of drainage problems…*' (*pauses*) '*…No, no, I see now*'.

SE: '*So if we cap these piles here...*' (*indicates several points on the sketch*)

Design co-ordinator: '*Yeah. OK. Let's do that*'.

The example shows how little was understood about the problem before commencing the design process. The problem itself (generating an effective way to provide structural support) arises out of the comparison of artefacts (sketch and drawing). Having recognised the problem, the two later went on to generate a solution. This solution is very different to the structured ways that the chart was used in Hutchins' description of navigation. Here, the structure of the problem in the work of construction is not fully specified, and the problem solvers must endeavour to clarify what they need to do to achieve their goal (in the case of the last example, to provide an appropriate form of structural support). Typical cockpit and navigational examples in the DCog literature do not involve this form of behaviour: the problem solvers already have a known problem and a well-practised set of procedures with which to generate a solution. The importance of this difference in problem solving for DCog in construction is that the problem space constantly changes over the time period, and repetition of activities is consequently infrequent.

The fieldwork also shows a substantial part of collaborative problem solving involves the system self-organising itself. While self-organisation is recognised and discussed by Hutchins as a feature of activity within functional systems in navigation (e.g., Hutchins 1991), we have observed extensive ongoing self-organisation in our research into loosely coupled systems. A crucial component of self-organisation is the agents' awareness about the state of the functional system, so they are able act in an appropriate and timely way to the ongoing situation. An analysis therefore needs to clearly show how agents manage to achieve situation awareness. However, situation awareness is not normally assumed to be a part of problem solving in DCog (because it is not always directly associated with a representation transforming activity) and as such has been largely ignored. We argue here, that in DCog studies of loosely coupled systems, situation awareness *must* be considered as a core feature of computational activity, even when it is not directly associated with *a particular* problem-solving event.

Organisational Structure and Problem Solving Dynamics In this section we show examples of how the components of the functional system were organised in the construction project. In the terms used by Hutchins (1995a), there were a number of documents that determined a 'standard operating procedure' (or SOP) including manuals and handbooks. However, their application was not as heavily enforced as navigational or aircraft cockpit systems, for a range of reasons (in navigation, SOPs control time criticality, safety, number of participants, and personal accountability). In construction, the organisational structure inherent in the SOP provided a basic structure, but allowed flexible interpretation by the actors involved. The example below shows this in an instance of how materials were ordered by the construction team and the resourceful way in which they managed to accomplish this (from fieldwork). To set the scene, a site engineer was discussing a concrete pour with a remote person over the telephone (note: only one side of the telephone conversation could be monitored by the fieldworker):

Site engineer: (*stands up and speaks loudly into telephone*) '*So, what I'm asking is*: *should we put concrete into the tower?*' (*raises his head and looks at the senior engineer with raised eyebrows*)

Senior engineer: '*Yes*'.

(*Site engineer, completes the telephone call, then lifts a radio to speak to a foreman to give the go ahead. A graduate engineer overhears this*:)

Graduate engineer: (*orients towards senior engineer*) *'Do you have any spare...* (*pause*) *... can I have three cubic metres*?'

Senior engineer: '*OK. Yeah.*'

(*Site engineer overhears this and radios through to the foreman to arrange it*).

In this observation, the open-plan office space allowed the overhearing of telephone conversations, and was used by the site engineer as a means of asking the senior engineer if he could go ahead with construction. This was not pre-planned, but arose from a request for information arising from a distant third party. A graduate engineer, in turn, overheard this, and made a request for materials, which was arranged by the site engineer. This saved money (and effort) by ordering one and not two separate concrete deliveries, yet none of this was planned in advance. This situation was able to take place because of the open-plan structure of the office space, but also because the participants knew that they did not have to order material through the specified SOP by ordering each load of concrete through a formal materials order. In this case, a separate order of concrete would have had to be made, tying up resources and losing the economies of scale that come with a large order.

A second example shows how the organisational structure acted as a *resource* for problem solving, while the mechanisms used in resolving the problem were socially mediated and negotiated between people. In the construction project, tasks were allocated to people through a number of mechanisms, dependant on hierarchical structures of seniority and the contextually dependent features of the setting. While allowing a degree of autonomy, the participants understood their responsibilities and the roles that they were expected to perform. The example below illustrates how knowledge distribution occurred through a variety of agents and media:

A graduate engineer was asked to check on the particular characteristics of a concrete mould (known as 'shuttering') by the clerk of works (who was employed by the RE). According to company regulations, queries raised by the RE or their staff should involve recording the problem, finding the answer, and filling out a 'works record', which would be sent to the site office, placed on file, and a copy sent on to the RE. Accordingly, the graduate engineer filled out a works record form with the problem request and sketched a diagram of the concrete shuttering and the setting it was placed in. He telephoned (someone) off-site, and discovered that the information he needed about using the shuttering was in an advertising/promotional leaflet sent out by the shuttering company, and which had just been sent on to be held on file in the team's office. Almost immediately, he noticed that this leaflet was lying on one of the foremen's desks, as he had been looking through it with an eye to ordering more materials. The engineer read off the technical details from a table on the leaflet and added this information to the form. He then posted the works record to the site office for circulation. As the works record was an official form, no accompanying contextual information was required because the nature of this structured document meant that it would always be processed in the same way. Due to the slow speed of the internal postal service, the engineer later went back on site, located the clerk of works and reported his findings personally.

The example demonstrates how the members of the local functional system (in this case, the graduate engineer, unknown telephone informer, foreman, clerk of works, and resident engineer) created and used representation-bearing artefacts 'on the fly' (the work record, the teams' file, sketch and leaflet) and over different channels of communication (spoken, post, and telephone). This process was not specified in the operating procedures laid out by the construction company and was generated and interpreted creatively by the participants. In this way, the SOP functioned as an (incomplete) organising resource rather than a rigid set of instructions, and it was loosely applied in the performance of work. It did not determine the physical actions required, which were selected according to a range of social, material and spatial factors. In this respect, it was creatively interpreted, rather than followed.

What is noticeable about the example is the way that the task involves both formal (i.e., established) work practices, some of which are given in the SOP, and an *ad hoc* approach to collaboration. Showing that work involves formal, or explicit, and informal, or tacit, features is not itself a novel finding (e.g., Grudin 1994). Nonetheless, this activity differs significantly from the practices noted in previous DCog analyses of work, in which the formal characteristics of work are exaggerated because of the particularities of the situations examined. Recognising the unpredictability of work and the agents' use of *ad hoc* work practices to deal with this is a critical feature of examining activity in the analysis of loosely coupled organisational systems. This interrelationship between the formal and *ad hoc* practices is important in understanding activity. While Hutchins (1995a, 1995b) and to an even greater extent Wright et al. (2000) place emphasis on the formal or proceduralised aspects of work, they do not ignore the informal aspect of collaboration entirely. However, within loosely coupled systems, the participants' ongoing orientation to the constantly shifting problem situation is central to their performance. The participants' access to a wide range of resources and the flexibility in the management of their own organisation means that agents within a functional system may exhibit many unique, situation-specific solutions to the problems they face. Application of a DCog approach in these settings must reflect this possible lack of precedents and the agents' artful use of the resources at hand.

Cycle Duration The cycle duration on the construction and design of the temporary works was highly variable—in terms of the project as a whole, the work was expected to take three years; far longer than the brief cycles of navigational fix taking. However, within this project, a number of smaller problem solving systems could be identified, each of which could be examined as a functional system in its own right. These different classes of problem ranged from specific temporary works designs (such as scaffolding towers), permanent works designs (ranging from pile placement to whole bridge designs), individual concrete pours, and so on, taking place over widely differing time-scales. Even similar tasks could take place over radically differing time-scales, depending on the availability of information about the problem, the intellectual problem solving resources, and the physical resources available to the functional system. Examples of this are impossible to demonstrate with snippets of observational data, and it is perhaps unnecessary to attempt to do so. The three-year example of the project is itself an ample demonstration of cycle duration in a design and construction project; work took place over an extended period, punctuated by periods of inactivity and intensive action. Of course, not all loosely coupled systems will periods of cycle duration as long as that of the construction project described, and the duration may also vary depending on the boundary that the analyst places on the frame of the analysis. For example, investigating the entire building process might typically take several years, but to examine the initiation of plans for the building project it may be feasible to observe only for a few months, or to examine the design construction of formwork, a few weeks may suffice.

For the cognitive ethnographer, the long duration of a project presents something of an obstacle: the complete design and construction project took far longer than the time available to study it. A three year long project is by no means unusual in the construction industry, and similarly long project durations are relatively common in industry in general. It is therefore unlikely that a study could be made of the process as a whole within even a medium term research project. This must be contrasted with the observations and analysis of navigational fix cycle or cockpit behaviour, which could be measured in seconds or minutes. While similarities exist in the abstract nature of problem solving itself, the practicalities of this difference between the two kinds of problems and settings could not be more different. The implication of this for a DCog analysis is that the cognitive ethnography cannot be a complete study of the work process, but only a part of it. Data will have to be collected from before the period of study, and the analyst will have to envisage how it will be continued following the field study, by projecting their findings forward. In comparison to the previous studies of DCog, in many cases there may be little chance of looking at another problem cycle.

Summary The examples presented in this study clearly demonstrate some of the differences with DCog in loosely coupled systems compared to ship navigation, aircraft cockpits and the other more limited situations in which we have seen problem solving examined through the use of DCog. Most noticeably, the work is heavily contingent, and the participants make extensive use of the wide range of representational resources that they find around them. The nature of this work was that it was highly context-dependent and unpredictable. While the engineers made plans, and organised labour and materials in an attempt to control the situation (the 'planned' component of activity), they were also constrained by the context within which the activity occurred. This involved adjusting their ongoing behaviours to the evolving circumstances on the site and making use of the resources available. This is not to

Table 9.1 A comparison of key dimensions in workplace settings

Key dimensions	Tightly coupled systems	Loosely coupled systems
Access to resources	Agents and representational artefacts are restricted to a predetermined set.	Agents and representational artefacts are unrestricted to a predetermined set and may change over time.
Problem structure	Well-structured, identifiable and expected problems that are recurrent.	A tendency toward ill-structured problems that have a high degree of uniqueness.
Organisational structure and problem dynamics	Organisation has pre-specified modes of operation, characteristic of tightly constrained and managed organisations with rigid modes of operation. Division of labour is well understood and 'standard operating procedures' underpin much of normal work. Problem dynamics are relatively stable over time.	Organisation's operation is only partially pre-determined; established work processes operate at an abstract level and are augmented by ad-hoc approaches in interpreting these high-level operational directives. Divisions of labour are informally defined and enforced. Problem dynamics are unstable and dynamic over time.
Cycle duration	Relatively short cycle for problem solving, coupled tightly to the task.	Problem-solving cycle tends to be variable, with similar classes of problem taking place over widely different time scales.

say that previous studies of DCog do not account for this contingency, but that it is more exaggerated in loosely coupled organisational systems and requires a greater degree of attention.

A Comparison of Study Settings

In assessing the analytic value of DCog applied to loosely coupled systems, it will be helpful to provide a comparison of their key constituent dimensions against settings in previously published studies of DCog (see Table 9.1 for a summary). In the instance of tightly coupled systems, we illustrate this with reference to navigation as this is broadly representative of the systems to which DCog analyses have been 'applied'. It is important to recognise that this comparison is illustrative, and it is not suggested that *all* systems of one kind or another will share *all* of these characteristics. Rather, these dimensions are intended to highlight areas where there are likely to be contextual differences which impact on our ability to employ DCog effectively. Please note that these dimensions are not completely distinct and there are interdependencies between them.

Access to Resources The main difference between tightly and loosely coupled organisational systems can be most clearly seen in terms of the distinction between their access to resources. In navigation, the system is closed: the process has a fixed and restricted set of resources and external agents are not able to involve themselves in the system. In loosely coupled systems, participants may call on a larger set of resources that might not be initially specified or known to be available at the beginning of the activity.

Problem Structure The problems that the two types of system have to solve may also be structured in different ways. In navigation the problem is 'well-structured' prior to its solution; the task is repetitive and agents are well practised in performing the task. In loosely coupled organisational systems, the problem is more usually 'ill-structured' and only becomes well-structured in performance as the agents learn about the problem during problem solving.

Organisational Structure and Problem Solving Dynamics The methods that are used to organise the co-ordination of activity differ between tightly and loosely coupled organisational systems. In navigation, the communication pathways are (necessarily) well specified and constrained to a number of pre-defined modes of operation. These are enforced by (naval) regulations, which prescribe the division of labour on particular tasks. However, in loosely coupled organisational systems, not all of the communication pathways are well-specified prior to problem solving, and their organisation are likely to be only partially constrained by pre-determined modes of operation. While we recognise that navigational work is not always routine, this is far more exaggerated in loosely coupled systems. There may be few absolute organisational structures, and the artefacts, communication pathways and participants available are likely to change over time. Some processes may be formally specified, but many are generated in an *ad hoc* fashion. Formal specifications may be stipulated at a high level, but the mechanics of implementing these are frequently left to the interpretation of participants, subject to professional and legislative constraints. In loosely coupled systems, procedures can be defined at an abstract level, and it may be left to the interpretation of individuals to decide on what actions to take as circumstances change.

The changing nature of the problems faced by the navigators and by agents in loosely coupled organisational systems differs substantially, and this has implications for the way that problem solving strategies develop and enter the culture of the workplace. Navigation by triangulation is an unchanging process, developed over centuries of practice. The procedures can remain unchanged over multiple fix cycles, and although each cycle may be of short duration in itself, they are highly repetitive. In loosely coupled organisational systems, the duration of the activities that they perform is likely to be far longer—for example, contract negotiations, design development, or product testing are lengthy activities, and over time, and even within the activity, procedures are likely to develop and adapt. This is unlikely to occur *within* the fix cycle, although Hutchins shows how the development of practice and of the practitioners (1995a, p. 372) does change over time—but over multiple cycles.

Cycle Duration The duration of the activity cycle differs substantially between tightly and loosely coupled systems. For example, in navigation, the 'fix cycle' is of short duration (a matter of minutes or seconds). These fix cycles are 'snapshots' in time, and each involves taking a bearing of the vessel's present location. However, in loosely coupled systems, activities can occur over much longer time-spans. In one sense, this reflects the difference in the work being done, where the tasks in the navigational system are more self-contained and time-critical.

These brief descriptions suggest that work performed by tightly and loosely coupled systems may be very different, in terms of their goals, technical resources and contexts of use. However, both may be seen as information processing systems with a similar high level (cognitive) structure and it is at this level that a common approach like DCog can be applied. Nevertheless, because these systems differ in significant ways, it is likely that the methods used to examine them will also have to differ. This suggests that we cannot draw directly from previous studies of DCog and apply them directly to the analysis of other workplaces. Rather, we need to consider the differences more closely, undertaking and drawing on new studies of organisational work to help us evolve an effective analytic approach for distributed cognition into these contexts.

Reflections on the Use of DCog

The fieldwork vignettes presented above demonstrate how the computational architecture of the functional system arose through the relationships between the agents, where the transformation of problem situation into design solution involved a variety of computations. This was implemented within a socially distributed cognitive architecture that incorporated a number of actors with different skills and roles, in combination with a range of other representational media, and operating in an environment rich in resources that structured these transformations. The social and organisational co-ordinating mechanisms that brought representations together worked in concert with the physical resources and constraints of the setting to determine the outcomes of these computations.

The descriptions[1] of action in the earlier sections focus on the informational transformations that take place, as representations are re-represented in various media. In most respects this is identical to the traditional form of DCog analysis. However, in this study, snapshots of activity were collected and presented from across a very large and distributed functional system, and through time. It is not the theoretical framework of DCog that differs in this then—the analysis is still of representations and processes, transformations and co-ordinations that allow aspects of the information processing structure of the functional system to be made explicit—but it does differ significantly in its practical application (limited by a small and

[1] In ethnography, this is also confusingly known as its *representation*. Use of the word 'representation' is distinct in meaning to its use in DCog, where it is applied to the symbol processing activities conducted in the functional system, and not by the analyst.

incomplete data set). In this sense, the description is less like a task analysis, in which a single episode is examined in detail (as with traditional DCog), and is more descriptive in the traditional sense of an ethnography, where the analyst selects representative aspects of the situation. However, what is being described is how representational transformations operate within these representative samples and extend into computations across the larger functional system.

In a study of a large and complex organisational system, such as the construction example given in this paper, data collection and its analysis will need to encompass the whole range of information processing activity that the functional system is capable of performing. As researchers examining these systems, we cannot fully specify the structure of the functional system. In the terms of cognitive science, the external symbol system derived will be incompletely specified and it cannot give us the granularity that Hutchins (1995a, 1995b) provides us with in a formal, functional structure of the activity. In practice, the wide scope of the organisational systems analysis means that the 'density' of the data collected and the coverage of the analytic methods and approach is far lower than the standard DCog approach can provide us with. In a loosely coupled activity system, we may have no means of examining all of the transformational actions undertaken. What an analysis can demonstrate is how, *in the situations observed*, the resources were applied to perform representational transformations achieving the system goal (or possibly failing to do so). Necessarily, some (possibly important) situations will not be observed (or be otherwise accessible) that are relevant. A realistic approach is therefore to do one of three things:

(i) We can reduce the activity examined to a subset of the complete activity. This is the kind of approach used successfully by DCog researchers to date. However, such an approach means that we cannot see how the activity within the wider system is performed.
(ii) We can accept that the level of granularity in the analysis will be lower, and that the descriptions of the actions performed will be less detailed. The analysis will explore the higher-level, organisational features of work rather than its practice. However, this approach means that we cannot look at the actual *use* of the representations and processes in detail. In many instances, this is what management and organisational analysts do, and in so doing, miss out the valuable roles that local practice and artefacts play in the performance of work.
(iii) We can focus on the significant actions that the participants and the ethnographer deem to be of particular importance to the performance of the functional system as a whole—whether it is particularly difficult, important to their co-ordination, where they have particular problems, or have to perform repair functions (i.e., resolve breakdowns) in particular situations.

This third option is the one that we would advise, and have used ourselves in the fieldwork described in this paper. This allows a degree of scope for the analyst to select aspects of the functional system that are particularly significant to the participants, and to do a deep analysis of the representational structures involved. However, this will not provide a complete description of actions in the functional system, and may be prone to a degree of subjective bias in the actions selected for detailed analysis. It relies to a greater extent on the importance attributed to situations by the ethnographer and participants than the standard DCog approach, but this is generally an accepted component of interpretive research (cf. Van Maanen 1988; Hammersley and Atkinson 1995) and is not, by itself, a failing. The method of data collection—ethnography—and the analytic framework suggested provides a means of examining *aspects* of information processing. It cannot be treated as a total description of an activity, but is a means of getting a deeper understanding about that activity within its context. We further suggest several features in the analysis of loosely coupled systems that differ from the traditional approach used by Hutchins and other DCog theorists:

(1) DCog is useful in illustrating how and which resources are used, but field data can only be understood as an *instance* of problem solving activity—at another time, activities may be performed very differently because the functional system has access to a wide range of resources. It should not be presented as a complete and systematic computational description of an activity in the way that Hutchins describes navigation or cockpit activity, because of the one-off uniqueness of the activity sequences observed.
(2) Piecing together the component parts of the analysis is not a simple matter. Because of the size and complexity of the activity system being studied, the analyst may have to link a large number of interacting local functional systems (for example, the generation of the design brief or a materials order) together to form a picture of higher-level problem solving (in this case, temporary works design). When doing the analysis, the cognitive ethnographer will need to identify which local functional systems form the 'trajectory of work' (Strauss 1985), and to investigate these with an aim to linking them together in the final analysis of the larger functional system.
(3) Cognitive ethnography cannot always present a complete description of computational activity, but only a part of it (over time, space and tasks) that is extrapolated to cover the whole problem solving cycle (a related point to 1 and 2 above). The analyst is unlikely to be able to examine another problem cycle to see how differences occur. However, it may be possible to study another organisation or activity system to compare activities. This should not be an attempt to validate the data through triangulation, but to present a cross-cultural dataset: analysts should not attempt to produce absolute answers to questions about the organisation of activity, but to provide a *better understanding* about the social practices and organisation of the functional system in context.
(4) The nature of the problem in the situation studied may be ill-structured and, as a consequence, the participants will strive to understand the nature of this problem and their co-workers' orientation to the problem and each other. The focus of analysis is therefore as much on understanding ongoing system's self-organisation as it is with directly looking at problem solving on the task. The analyst will need to pay increased attention to how agents in the functional system are *made aware* of ongoing, but problem-unrelated, situation monitoring so that they can self-organise. This is related to the agent's situation awareness

which, we argue, must be considered as a core feature of activity, even when it is not directly associated with a specific problem-solving event. Hutchins does discuss situation awareness (1995a), but only in relation to known environmental events: for example, when landmarks come into view, when observations are shouted down the intercom, when marks are made onto tables and charts or when agents within the functional system leave or join.

Conclusion

The framework of DCog allows the analyst to examine the computational nature of an activity, and the problem solving activity that occurs over a distributed group of individuals and artefacts. The analysis lies not in the abstract processes of an activity, or in a description of the communications that made it possible, but on how the activity and its co-ordination are inter-linked through task performance. In this, we do not seek to radically alter the foundations of the DCog approach: DCog is a broad church and does not advocate adherence to a particular method or approach to gathering data about the functional system (Hutchins 1995a; Hollan et al. 2001). It is important to recognise that we are not critical of the approach used in traditional DCog; rather we show how DCog can be applied to situations in which the traditional DCog analytic *approach* (exemplified by Hutchins 1995a) would not be possible or appropriate.

In some ways, what we are proposing in terms of the analysis of open and loosely coupled systems is not dissimilar to that commonly performed in ethnomethodological studies of action (Garfinkel 1967; Heritage 1984). The scale of the analytic frame, the form of the data collected, and the ways that actors and artefacts may be enrolled into systems in an ad hoc way can be seen as similar. Yet these approaches are quite distinct as theoretical approaches and in what they offer in an analysis. The idea that a system might be 'cognitive' in any way is something of an anathema to ethnomethodologists (Heritage 1984), who take a distinct different perspective to cognitive science (indeed, they are explicitly atheoretical in their approaches to examining human action) and hold well-articulated arguments against cognitivism and the idea of goal-directed approaches to action.

A core difference between the sorts of distributed cognitive analysis performed in this study and the sort of study performed in tightly coupled systems is that we do not attempt to describe a complete computational structure for the functional system. Rather, we look to the framework of DCog as a means of exposing the more highly relevant information processing aspects of the activity (as judged by the fieldworker or workplace analyst). This is in contrast to much of the existing work in DCog, which attempts to produce well-defined accounts of functional systems that capture most, if not all, of the information processing characteristics inherent in them.

While the form of analysis advocated here for loosely coupled systems cannot be as precise as that produced by an analyst working with more tightly coupled systems, the approach is clearly of relevance to organisational theorists and systems designers. It changes the remit of the analyst from providing a near complete specification of a problem solving activity into one that demonstrates how to focus the analysis of ethnographic data onto the informational component of an activity. For applied researchers, this focus on information processing to show how representations and processes can be appropriated for use in performing and co-ordinating distributed problem solving activity has direct implications for systems design: it foregrounds the role of information and unites the social, organisational, psychological, situational, and artefactual elements that contribute to problem solving. The value of this is that analysts are able to see the interconnectedness of these factors in the activity, and to envisage how any interventions might have beneficial or harmful impacts on the activity system's information processing characteristics.

Acknowledgements The article has greatly benefited from discussions with Jon Hindmarsh, Kenton O'Hara and Barry Brown.

References

Alač, M., & Hutchins, E. (2004). I see what you are saying: action as cognition in fMRI brain mapping practice. *Journal of Cognition and Culture*, *4*, 629–661.

Artman, H., & Wærn, Y. (1999). Distributed cognition in an emergency co-ordination center. *Cognition Technology & Work*, *1*, 237–246.

Blanford, A., & Furniss, D. (2006). DiCoT: a methodology for applying distributed cognition to the design of teamworking systems. In S. Gilroy & M. Harrison (Eds.), *LNCS: Vol. 3941. DSV-IS'05 (International workshop on design, specification and verification of interactive systems)* (pp. 26–38). Berlin: Springer.

Engestrom, Y., & Middleton, D. (1996). *Cognition and communication at work: distributed cognition in the workplace*. Cambridge: Cambridge University Press.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs: Prentice-Hall.

Giere, R. N. (2004). The problem of agency in scientific distributed cognitive systems. *Journal of Cognition and Culture*, *4*, 759–774.

Grudin, J. (1994). Eight challenges for developers. *Communications of the ACM*, *37*, 93–104.

Halverson, C. A. (1995). *Inside the cognitive workplace: new technology and air traffic control*. PhD thesis, Department of Cognitive Science, University of California, San Diego.

Halverson, C. A. (2002). Activity theory and distributed cognition: or what does CSCW need to do with theories? *Computer Supported Cooperative Work*, *11*, 243–267.

Hammersley, M., & Atkinson, P. (1995). *Ethnography: principles in practice* (2nd ed.). London: Routledge.

Hazelhurst, B. L. (1994). *Fishing for cognition: an ethnography of fishing practice in a community on the west coast of Sweden*. PhD thesis, Department of Cognitive Science, University of California, San Diego.

Heath, C., & Luff, P. (1991). Collaborative activity and technological design: task coordination in London Underground control rooms. In *Proceedings of the 2nd European conference on computer supported cooperative work*, Amsterdam, The Netherlands (pp. 65–80).

Heritage, J. (1984). *Garfinkel and ethnomethodology*. Cambridge: Polity Press.

Holder, B. E. (1999). *Cognition in flight: understanding cockpits as cognitive systems*. Dissertation, Cognitive Science, University of California, San Diego.

Hollan, J. D., Hutchins, E., & Kirsh, D. (2001). Distributed cognition: a new foundation for human-computer interaction. *ACM Transactions on Computer-Human Interaction*, *7*, 174–196.

Hughes, J. A., Randall, D., & Shapiro, D. (1992). Faltering from ethnography to design. In *CSCW 92: sharing perspectives. Proceedings of the conference on computer supported cooperative work* (pp. 115–122). New York: ACM Press.

Hutchins, E. (1991). Organizing work by adaptation. *Organization Science*, *2*, 14–39.

Hutchins, E. (1995a). *Cognition in the wild*. Bradford: MIT Press.

Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science*, *19*, 265–288.

Hutchins, E., & Klausen, T. (1991). Distributed cognition in an airline cockpit. In Y. Engeström & D. Middleton (Eds.), *Communication and cognition at work* (pp. 15–34). Cambridge: Cambridge University Press.

McGarry, B. (2005). *Things to think with: understanding interactions with artefacts in engineering design*. PhD Dissertation, University of Queensland, St Lucia, Brisbane, Australia.

Murray, D. (1993). An ethnographic study of graphic designers. In *Proceedings of the third European conference on computer-supported cooperative work* (pp. 295–309). Netherlands: Kluwer.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Norman, D. A. (1993). *Things that make us smart*. Reading: Addison-Wesley.

Perry, M. (2003). Distributed Cognition. In J. M. Carroll (Ed.), *HCI models, theories, and frameworks: toward an interdisciplinary science* (pp. 193–223). San Francisco: Morgan Kaufmann.

Resnick, L. B., Levine, J. M., & Teasley, S. D. (1991). *Perspectives on socially shared cognition*. Hyattsville: American Psychological Association.

Rogers, Y., & Ellis, J. (1994). Distributed cognition: an alternative framework for analysing and explaining collaborative working. *Journal of Information Technology*, *9*, 119–128.

Salomon, G. (1993). No distribution without individual's cognition: a dynamic interactional view. In G. Salomon (Ed.), *Distributed cognitions* (pp. 111–138). Cambridge: Cambridge University Press.

Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence*, *4*, 181–204.

Simon, H. A. (1981). *The sciences of the artificial* (2nd ed.). Cambridge: MIT Press.

Strauss, A. (1985). Work and the division of labour. *The Sociological Quarterly*, *26*, 1–19.

Van Maanen, J. (1988). *Tales of the field: on writing ethnography*. Chicago: University of Chicago Press.

Wright, P. C., Fields, R. E., & Richardson, M. D. (2000). Analyzing human-computer interaction as distributed cognition: the resources model. *Human-Computer Interaction*, *15*, 1–41.

Chapter 10
Thinking with External Representations

David Kirsh

Abstract Why do people create extra representations to help them make sense of situations, diagrams, illustrations, instructions and problems? The obvious explanation—external representations save internal memory and computation—is only part of the story. I discuss seven ways external representations enhance cognitive power: they provide a structure that can serve as a shareable object of thought; they create persistent referents; they facilitate re-representation; they are often a more natural representation of structure than mental representations; they facilitate the computation of more explicit encoding of information; they enable the construction of arbitrarily complex structure; and they lower the cost of controlling thought—they help coordinate thought. Jointly, these functions allow people to think more powerfully with external representations than without. They allow us to think the previously unthinkable.

This chapter is an inquiry into why thinking and sense making, so often, is interactive. By 'interactive' I mean a back and forth process: a person alters the outside world, the changed world alters the person, and the dynamic continues. Reading a text silently is not an interactive process, for my purposes here, though it is extremely active. Reading and underlining the text, or reading and summarizing it, even reading and moving one's lips, are.

The puzzle that interaction raises about sense making and thinking can be posed like this. In a closed world, consisting of a person and an external representation—a diagram, illustration, spoken instruction, or written problem statement—why do people do more than just think in their heads? If we assume there is no one to ask, no tool to generate novel results, no clock to provide chronometric input, no process to run and observe, then there is nothing external, no oracle or tool, that a person can consult or manipulate, that yields new information. The environment contains nothing that could not be inferred through reflection, at least in principle. So why

D. Kirsh (✉)
Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92093-0515, USA
e-mail: kirsh@ucsd.edu

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_10,

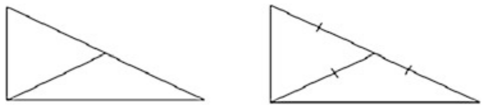


Fig. 10.1 By drawing an example of a right angle triangle and median, it is easier to understand the claim 'in a right-angled triangle the median of the hypotenuse is equal in length to half the hypotenuse'. The illustration does not carry the generality of the linguistic claim, but it is easier to convince ourselves of its truth. In 10.1**b** the equalities are explicitly marked and the claim is even easier to read, and helps hint at problem solving approaches

bother to make marks, gesture, point, mutter, manipulate inert representation, write notes, annotate, rearrange things, and so on? Why not just sit still and 'think'?

Figure 10.1a illustrates a simple case where interaction is likely. A subject is given the sentence, S_1:

> *A basic property of right-angled triangles is that the length of a median extending from the right angle to the hypotenuse is itself one half the length of the hypotenuse.*

What do people do to understand S_1? After re-reading it a few times, if they have a good imagination and some knowledge of geometry, they just think. They make sense of it without physically interacting with anything external. Most of us, though, reach for pencil and paper to sketch a simple diagram to help, such as Fig. 10.1a or 10.1b. Why? If the sentence were "The soup is boiling over" or "A square measuring 4 inches by 4 inches is larger than one measuring 3 inches by 3 inches" virtually no on would bother. Comprehension would be automatic.

Anyone who believes in situated, distributed, or extended cognition will have a ready explanation. Cognitive processes flow to wherever it is cheaper to perform them. The human 'cognitive operating system' extends to states, structures, and processes outside the mind and body. If it is easier to understand a sentence by creating a diagram to help interpret it, then one does that instead of thinking internally alone. The analogy is with a computer system that has memory systems and scratch pads in different media and locations. The decision whether to work out a computation in one or more scratch pads is determined by the type of operators available in each, the cost of operating in each pad, and the availability of space to work in. Processes should migrate to wherever they are best, or most easily, performed.

Figure 10.2 is suggestive of this view of extended or distributed cognition. Because people are embedded in their environments, they are densely coupled to the outside. Cognitive processes drift to wherever they are more cost effective. It's all about the cost structure of computation in each of the interconnected sub-systems. Evidently, when pen and paper is handy, and when the sentence is complex enough, it pays to make a good illustration; it reduces the overall cognitive cost of sense making.

Although I believe this is, essentially, a correct account, it is only one of the reasons people interact with external representations. The others have to do with ways

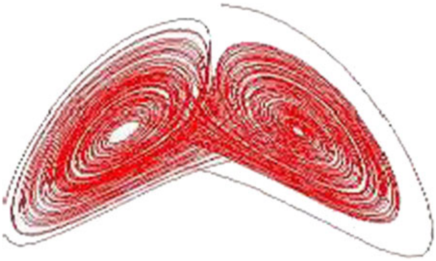


Fig. 10.2 This image of a coupled system represents the state space trajectory over time of certain cognitive processes. Processes readily move from one side to the other, wherever the cost of an operation is lower

in which changing the terrain of cognition can do more than change cost structure. Chiefly, these involve access to new operators—you can do something outside that you cannot inside; or, you can encode structures of greater complexity than you can inside, external mechanisms allow us to bootstrap to new ideas and new ways of manipulating ideas; or, thirdly, you can run a process with greater precision, faster, and longer outside than inside—you can harness the world to simulate processes that you cannot simulate internally, or cannot simulate as well. In short, these other ways concern changing the domain and range of cognition. This is a striking claim I will justify toward the end.

There is a further reason people interact with external representations: to prepare themselves to coordinate internal and external states, structures, and processes. This feature of interaction is fundamental to our understanding of external representations but rarely studied (see Kirsh 2009a, 2009c). For example, before subjects use a map to wayfind, they typically orient or 'register' the map with their surroundings; they put it into a usable correspondence with the world. Many people also gesture, point, talk aloud, and so on. In principle, none of these actions are necessary to establish a correspondence between elements in the map and the things those elements refer to. Eye movements, mental projection, and other non-interactive techniques may suffice for map-based navigation. But external interactions are commonplace, and a major aspect of understanding representations.

I have found these 'extra' actions also pervasive when people try to understand and follow instructions. In pilot studies, we found that subjects engage in 'interpreting' actions when they follow origami instructions. They register the origami paper with the instruction sheet; they point to elements on the instruction sheet and then focus attention on the counterpart property of the paper; they mutter, they gesture, they move the paper about. This activity is part of processing the meaning of the instructions.

To a lesser degree, the same thing often happens when non-expert cooks follow recipes. They keep place with their finger; they arrange the ingredients to encode their order of use (Kirsh 1995); they read the recipe aloud, ask themselves questions about ingredients, or mutter reminders. We observe similar behavior when people assemble furniture. Far from just thinking and then executing instructions, people perform all sorts of apparently 'superfluous' actions that facilitate comprehension. They point, mumble, move the instruction manual around, encode order of assembly in the arrangement of pieces. These actions are not incidental, they are often vitally important to sense making and effective action.

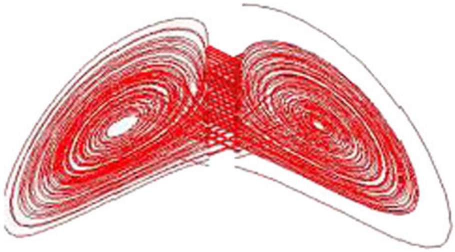


Fig. 10.3 This illustration suggests that there are three cost structures: the cost of inner operations on states, structures, processes, the cost of outer operations on states, structures, processes, and the cost of coordinating inner and outer processes, which includes the cost of anchoring projections, and the cost of controlling what to do, when, and where to do it

One function of these extra actions is to help people anchor their mental processes on external features or processes. Another is to help them tease out consequences, to deepen their semantic and pragmatic processing of the instructions. In both cases, people need to establish a coordination between what goes on inside their heads and what goes on outside. They construct a correspondence, a coordination relation, a synchronization. Because these coordination processes are not cost-free, Fig. 10.2 is overly simple. We have to add to the figure a representation of the coupling process itself, the special actions performed to establish a cognitive link. Figure 10.3 illustrates this added cost-laden process: anchoring (see Kirsh 2009b, for an initial discussion of this third cost space).

As important as the anchoring—or grounding process—is, I restrict my focus, in the remainder of this chapter, to ways we interact with representations to alter the cognitive terrain rather than the interactions we perform to prepare ourselves to engage the external part of that terrain through anchoring.

Materiality and Its Consequences

The argument others and I have long advanced is that people interact and create external structure when thinking because:

> Through interaction it is easier to process more *efficiently* and more *effectively* than by working inside the head alone (Clark 2008; Kirsh 1995, 1996, 2009a, 2009b, 2009c; Kirsh and Maglio 1994).

Efficiency usually translates as speed-accuracy. Interactive cognition regularly leads to fewer errors or to greater speed. Effectiveness usually translates as coping with harder problems. Interactive cognition regularly helps subjects to compute more deeply, more precisely, and often more broadly.

The idea is that by operating with external *material*, with pen, paper, ruler, and then working to meet one's goals and sub-goals using that external material—*draw* a triangle, *mark* the half point of the hypotenuse—subjects benefit from physical

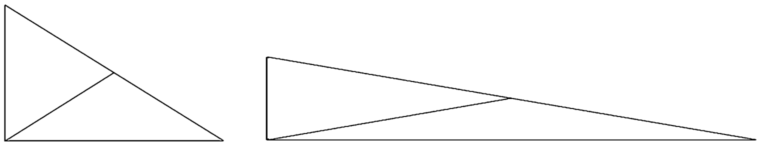


Fig. 10.4 Choices must be made when drawing a triangle. Should the triangle be long and short? Isosceles? Will any of these choices affect the truth of the sentence? By having to resolve these questions, subjects are helped in the problem solving process

constraint and visual hints that help cognition. This plays out in a few ways. For instance, the constructive process helps drive interpretation. Because action is primarily serial, it is incremental; a structure emerges step-by-step and a subject must resolve specific problems. What size should the base and height be? Does it matter? Does the median bisect the right angle? Working with tools and the external structure, moreover, grounds interpretation in an ever more constrained case study. After choosing the size of the right angle triangle, the requirement to split the hypotenuse in half is very concrete. It is now 'split this hypotenuse'. This incremental, interactive process, filled with prompts, hints, visible possibilities and impossibilities, provides more constraint than mentally computing a conceptual whole solely from the semantics of linguistic parts. The linguistic formulation is more general, but it is also less constrained (see Figs. 10.1 and 10.4).

A second way materiality figures in cognition is by explicitly involving visual and motor cortex. When a structure is viewable and drawable, its properties prime a constellation of associations. Just by grappling with external material—using rulers, making lines intersect—and then looking at the results, a set of properties and possibilities of forms are encountered and primed. For instance, if two lines intersect then they define a set of angles. It is natural for visual attention to focus on estimating angles. Are they equivalent? If the triangle has a right angle, then automatically a network of spatial concepts related to right triangles are activated, particularly associations derived from previous work with diagrams of right triangles. These visual and physical associations may be different and more extensive than associations derived from verbal accounts. This is apparent whenever a tool is in hand. Rulers prime measuring actions and thoughts; protractors encourage thoughts of angles and degrees.

The benefits of interacting with an external representation are especially clear for complex structures. As the complexity of a linguistic specification of a visual structure increases, it becomes more rewarding to make sense of the sentence by constructing a physical drawing and looking at it, than by constructing that geometric form in one's mind's eye and making sense of the sentence internally. Most people find it easier to think in terms of physical lines than in terms of the mental counterparts of lines, particularly the more lines there are, or the more complex the structure. Even though some people can do things in their heads that others cannot, there is always a point where internalist cognitive powers are overwhelmed and physical realization is advantageous (see Kirsh 2009b). Thus, although from a purely logical point of view, a closed system of world and person contains no additional information after that person has drawn an interpretation than before, there nonetheless are important changes wrought by interaction that can positively alter the cognitive terrain. Specifically, these interactive changes concern:

- What's *active inside* the person's head—what's being attended to, what's stored in visual or motor memory, and what's primed—an external structure encourages a visual scanpath that activates expectations, drawing the structure displays angles, lengths, and will cause distant cognitive associations in motor and visual cortex;
- What's *persistent outside*, and in the visual or tangible field—an external structure holds a structure constant until it is added to; the structure does not decay the way mental structures and processes do, and it supports repeated perceptual inquisition;
- How information is encoded, both inside and outside in virtue of interaction. Because there is an external structure present, subjects can try out different *internal and external representational forms*, the two forms can play off each other in an interactive manner, leading to new insights.

The upshot is that, often, humans are able to improve their thinking and comprehension by creating and using external representations and structures. By working outside, they change what is inside and interactively they can reach new thoughts. This may be stunningly obvious, yet it is sufficiently foundational and far-reaching to deserve analytic and empirical exploration.

Let me press this idea further by turning now to seven distinct benefits that externalization of structure confers.

Shareable and Identifiable Objects of Thought

When people externalize a structure, they are communicating with themselves, as well as making it possible for others to share with them a common focus. An externalized structure can be shared as an object of thought. This reification of internal object—this externalization—has benefits for both parties.

Here is an example. In Fig. 10.5, an explicit geometric form has been added to the body position of a dancer. Using a video to demonstrate torsion, Bill Forsythe, a noted choreographer, had his colleagues visually annotate key body features on the video as he spoke. He first identified points on his body, orally, then, as he turned his discussion to line segments, such as the line between elbow and hand, these were superimposed on the video, and finally he talked of joining the segments into a three-dimensional trapezoid, and his viewers saw a representation of the three-dimensional form appear on screen. It was then easy for viewers to see the effect of movement on deformations of the trapezoid. Forsythe relied on his listeners seeing the visible annotation, the trapezoidal structure, as he explained the ideas of torsion, sheer, and body axes (Forsythe 2008).

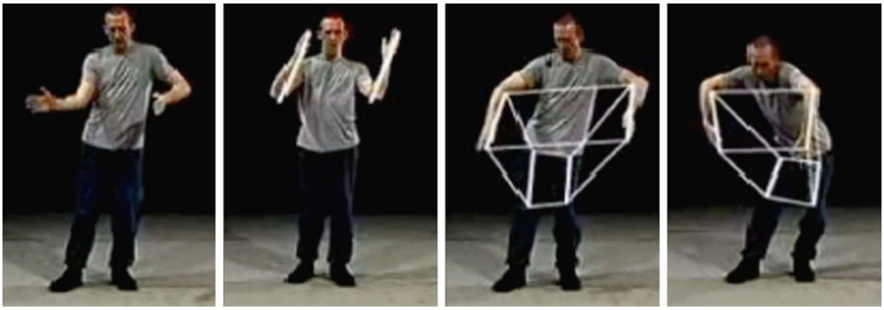


Fig. 10.5 Bill Forsythe, a noted contemporary choreographer, has begun documenting certain concepts and principles of choreography in film. Here he explains torsion. The annotation makes it easy for the audience to refer to otherwise invisible structures

One virtue of this particular annotation is that by having verbally defined the structure to be manipulated, and then visibly locating it on his body, the choreographer and anyone looking at the video, know that if they refer to any visible part of the trapezoid their reference will be understood. They can ask pointed questions about how the shape figures in what the speaker is saying, or even how some specific feature—the apex or base—figures in an abstract idea. For instance, once there are external lines and planes anyone can ask the speaker, or themselves, which body positions keep the volume of the shape constant, or which movements ensure the top plane remains parallel to the bottom plane. Choreographers find such questions helpful when thinking about body dynamics and when they want to communicate ideas of shearing and torsion to their dancers. But they are hard to understand if the group does not share a visual or projected image of a transforming shape.

Physically reifying a shape through annotation adds something more than just providing a shared reference; it provides a persistent element that can be measured and reliably identified and re-identified. Measurement is something one does after a line or structure has been identified. This need not always require an external presence. Some people are able to grasp the structure of a superimposed trapezoid purely by mentally projecting an invisible structure onto the body. They listen to the speaker; watch his gestures, and project. But even for these strong visualizers, annotating still helps because once something is externalized it has affordances that are not literally present when projected alone.

For instance, when the lines of a shape are externalized we can ask about the length of the segments and their angles of intersection. We know how to measure these elements using ruler and protractor. Lines afford measuring. Granted, it is still possible, though not easy, to measure the length of mentally projected lines if the subject is able to appropriately anchor the projected lines to visible points. A choreographer, for instance, can refer to the length of someone's forearm through language or gesturally mark a structure without having to annotate a video. But can he or she refer reliably to the length of lines *connecting* the top and bottom planes of a complex structure without having those planes visibly present? Those lines have to be anchored on the body. If a structure is as complex as a truncated pyramid, which has eight anchor points, it must be constructed in an orderly manner, much as Forsythe did in his annotated video, else there is too much to keep track of mentally. This does not decisively show that such structures cannot be identified and marked out by gesture and posture without visible annotation. But the complexity of mental imagery and mental projection goes way up as the number of anchors increases; or when the target body moves the anchor points; or, worst of all, if invisible anchor points are required, as would be the case if the conceptualized pyramid were to extend right up to its apex. The peak itself would be floating in air, unconnected to anything material, anchorless. Imagine trying to use that invisible anchor as an anchor for something else. By contrast, once the form is made manifest in visible lines, all such elements can be explicitly referred to, even visibly labeled; they can be located, measured, intentionally distorted if so desired, and the nature of their deformation over time can be considered. They become shared objects of thought.

This is worth elaborating. To say that something is, or could be, an object of thought implies the thinker can mentally refer to it—in some sense the thinker can grasp the referent. A shared object of thought means that different thinkers share mechanisms of reference and for agreeing on attributes of the referent. For instance, Quine (1960), following Strawson (1959), argued that objects must have identity conditions, as in his motto "No entity without identity". Entities have to be identifiable, re-identifiable, and individuatable from close cousins. Would the structures and annotations in Fig. 10.5 meet those criteria if imagined or projected mentally? It depends on how well they are anchored to physical attributes. Certainly there are some people—choreographers, dancers, and people with wonderful imaging abilities—who can hold clear ideas of projected structure, and use them to think with. As long as there is enough stability in the 'material anchors' (Hutchins 2005) and enough expertise among the subjects to ensure a robust projection, the lines and shapes these experts project onto the visible environment meet most criteria of 'entification', though, of course this is purely an empirical claim. But most of us find that it is easier to think about a structure that has been reified by adding visible or tangible elements to the environment. The structure is more vivid, more robust, clearer—a better object of thought. Almost everyone needs to see the lines and shapes to see subtle geometric relations between them. So, we create external structure. It is by this act of materializing our initial projections, by forming traces of those projections through action, or material change, that we create something that can serve as a stepping-stone for our next thoughts.

This interactive process of *projecting structure then materializing it*, is, in my opinion, one of the most fundamental processes of thought. When we interact with our environment for epistemic reasons, we often interact to create scaffolds for thought, thought supports we can lean on. But we also create external elements that can actually serve as vehicles for thoughts. We use them as things to think with.

All too often, the extraordinary value of externalization and interaction is reduced to a boring claim about external memory. "Isn't all this just about offloading memory?" This hugely downplays what is going on. Everyone knows it is useful to get things out of the head and put where they can be accessed easily any time. It is well known that by writing down inferences, or interim thoughts, we are relieved of the need to keep everything we generate active in memory. As long as the same information can be observed and retrieved outside, then externalizing thought and structure does indeed save us from tying up working memory and active referential memory.

But memory and perception are not the same thing. Treating information to be the same whether it is outside or inside ignores the medium specific nature of encoding. The current view in psychology is that when we visually perceive an external structure, the information that enters is stored first in visuo-spatial store (Baddeley 2000; Logie 1995), before being processed for use in later mental processes. Since the form a structure is encoded in profoundly affects how easily it can be used in a process, it is an open question how much internal processing is necessary to convert an external structure into an internal structure that is usable. Accordingly, it cannot be assumed, without argument, that the costs are always lower in perceptually retrieving information than 'internally' retrieving information, even if that information is complex and voluminous and something we would normally assume is more efficiently stored externally. The strength of this concern is obvious if the information element to be perceived is buried in visual clutter. Much will depend on visual complexity, the form information is encoded in, how easy it is to perceive the structure when it is wanted, and so on. Even when an object of thought is present in a clear and distinct way—as Forsythe's graphical annotations are—it still must be perceived, then gestalted, and conceptualized. Do we really know the relative cost of grasping an externally represented content versus an internally represented one?

The implication is that using the world as external storage may be less important as a pure source of cognitive power than using the world for external computation. Things in the world behave differently than things in the mind. For example, external representations are extended in space, not just in time. They can be operated on in different ways; they can be manually duplicated, and rearranged. They can be shared with other people. Tools can be applied to them. These differences between internal and external representations are incredibly significant. They are what makes interactivity so interesting.

I turn now to another of these differences: the possibility of manually reordering physical tokens of statements. Because of rearrangement, it is possible to discover aspects of meaning and significance—implications—that are hard to detect from an original statement when viewed in isolation. By reordering and rearranging what is close to what, we change a token's neighborhood, we change the space of what is cognitively near.

Rearrangement

The power of physical rearrangement, at least for vehicles of propositions, such as sentences, logical formulae, pictorial narratives, is that it lets us visually compare statements written later with those written earlier; it let's us manipulate what is beside what, making it easier to perceive semantically relevant relations. For instance,

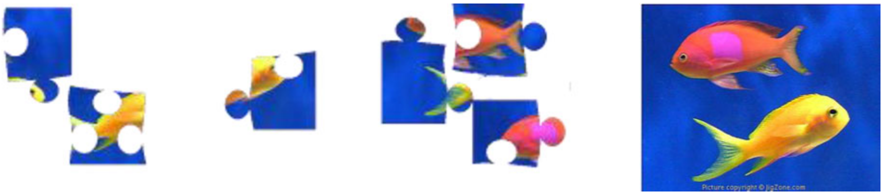


Fig. 10.6 Can the jigsaw images on the *left* be perfectly assembled into the picture on the *right*? If you could rearrange the pieces, the answer would be trivial. The answer is no. Can you see why? Why, in general, is it easier to solve jigsaw puzzles tangibly?

we can take lemmas that are non-local in inference space—inferences that are logically downstream from the givens and usually discovered later, hence written further down the page—and rewrite them so they are now close to earlier statements. Statements that are distant in logical space can be brought beside each other in physical space. If we then introduce abbreviations or definitions to stand in for clusters of statements, we can increase still further the range of statements we can visually relate. This process of inferring, duplicating, substituting, reformulating, rearranging and redefining, is the mechanism behind proofs, levels of abstraction, the Lisp programming language, and indeed symbolic computation more generally.

The power of *rearrangement* is shown in Fig. 10.6. The problem is to determine whether the six pieces on the left are sufficient to build the form on the right. What do you need to do to convince yourself? Since the problem is well posed and self contained, the question again, is 'why not just work things out in your mind?' In Fig. 10.6, you have no choice: because the pieces are not movable, no doubt, you will confine your thinking to looking and imagining the consequences of moving and rotating them. But, if the problem were posed more tangibly, as a jigsaw puzzle with movable tiles, wouldn't it be easier to try to *construct* an answer in the world than to think through an answer internally?

Reorganizing pieces in physical space makes it possible to examine relations that before were distant or visually complex (e.g., rotations and joins). By re-assembling the pieces, the decision is simply a matter of determining whether the pieces fit perfectly together. That is a question resolvable by physically fitting and visually checking. Interaction has thus converted the world from a place where internal computation was required to solve the problem to one where the relevant property can be perceived or physically discovered. Action and vision have been substituted for imagery, projection, and memory. Physical movement has replaced mental computation. Instead of imagining transformations, we execute them externally.

It is tempting to interpret the benefits of rearrangement entirely in cost structure terms: processes migrate to the world because they are cheaper or more reliable there. Evidently, physical manipulation, at times, is cognitively more efficient and effective than mental manipulation. So, on those occasions, it is rational to compute externally.

And sometimes that is all there is to it. For example, in Tetris, subjects can choose between rotating a tetrazoid in their heads and rotating it in the world (Kirsh and

Maglio 1994). Since physical rotation is, in fact, a bit faster than mental rotation, the cost incurred by occasionally over-rotating a piece in the world is more than made up for by the benefits that come from the faster and less error prone decision making based on vision.

Yet, it is not always so. In solving jigsaw puzzles, more is at stake than cost alone. As the descriptive complexity of the board state increases, there comes a point where it is hard, if not impossible, for someone to hold the complete structure in mind. The very act of trying out a move mentally causes the internally maintained structure to degrade. Imagine trying to assemble twenty separate pieces in your mind, and then checking to see if the twenty-first will fit anywhere in the mentally sustained assembly. The twenty-first piece may be the last straw, total overload, causing the whole mental structure to lose integrity.

The analogy is with swap space in a computer. Once a threshold of complexity is reached, a computer begins to degrade in its performance. In fact, if its flailing is serious enough, it reaches a standstill, where it takes so much of its working memory to hold the board state, that the simple act of changing that state exhausts memory. The system lacks the resources to keep track of what it has tried already and what remains to be tried. It has to place in long term memory the part of the board state it is not currently checking, so that it can process the steps in its program telling it what to do next. Then, to do the next thing, it has to bring back part of the board state in long-term memory, and swap out the control state. The result is that the system may cycle endlessly in the same subset of states, never canvassing the part of the state space where the solution is to be found. Zero progress.

It is not quite like that in the world. Because of physical persistence, the board remains the same before and after a subject thinks about moves. Unlike the mental realm, the stability of a physical state is not significantly affected by its complexity. A twenty-piece assemblage is just as stable as a ten-piece assemblage.

There are limits in the physical world too. Once a board arrangement has been changed physically, the previous state is lost, unless a further trace, an annotation was created, or a digital image taken. So searching for a solution in the world, as opposed to in the head, is not always better. But with enough externalization of state—enough external record keeping—there are jigsaw puzzles that can be solved physically that would be impossible to solve in the head, through mental simulation alone. We can push the complexity envelope arbitrarily far. This cannot be done in the head alone.

I will return to this topic of in-principle differences between mental and physical simulation at the end of the essay.

Physical Persistence and Independence

Both rearrangement and having stable objects to think with rely on physical things being persistent. The next key difference between internal and external representations, then, concerns the difference in their stability and persistence over time.

Rearrangement of jigsaw pieces is possible because the different pieces to be arranged are simultaneously present. If six pieces were present before rearrangement, there are six after. Pieces can be moved nearer to each other without destroying their integrity. Even though things are not quite as simple with thinking with physical tokens of sentences, we still can be confident that we have the same thought before and after moving a sentence. Because it is easy to detect differences in a sentence token simply by comparing the original with its copy—that is, before and after copying a sentence inscription—we depend on physical persistence to ensure we do not change the object of thought just by copying or moving tokens. Other things equal, the sentence 'this sentence has five words' means the same whether printed on the right or the left of the page, and whether printed yesterday or today.

The case is rather different for mental representations. How can a subject be sure that the mental image in mind at time t_1 is the same as the one at t_0? And how can a subject know whether the addition of another mental image, or a simple rotation of a mental image, has not changed the original image? The only reliable test is whether the image is caused by the same external structure on both occasions. If that structure is not present, there is no objective touchstone to decide sameness. There is just subjective belief. For Wittgenstein (1953) this was a source of skepticism concerning the possibility of knowing one's mental state without outside referents to ground it. No inner state or inner process without outer criterion. Hence, without external support, there might be no way of knowing whether one has the same thought on two occasions.

The brute fact of physical persistence, then, changes the reliability, the shareability, and the temporal dynamics of thinking. It is easier to have the same thought tomorrow, if the vehicles encoding the thought, or the cues stimulating the thought, are the same as today's. That's why writing helps. When the vehicle is external we can also count on other people ratifying that it remains the same over time. So, we can be confident that if we think we are reading the same sentence on two occasions, there is a fact of the matter. Similarly, we can be confident that if we interact with an external representation and we think we have left it unchanged, our judgments are more reliable than those concerning our beliefs about internal representations. Moreover, in the outside world, there is widespread empirical agreement on the effect of interaction—we know there is a broad class of transformations that leave structures invariant, for example: rotation, translation, lighting change, and so forth. There is no comparable principle for internal representations. We have no way of knowing the constancy of our inner life. This means that we have a better idea of the effect of interacting with external representations than with internal ones.

Physical persistence also differs from mental persistence, and transient mental presence, in increasing the *range of actions* a subject can perform on the underlying thing encoding the representation—the vehicle. In Fig. 10.5, for example, the truncated '3D' trapezoid is displayed as a line drawing on the choreographer's body. It is shown in stop action. Measurements can be made because the visible structure—the trapezoid—can be frozen for as long as it takes to perform the measurements. Tools can be deployed. The materiality of external representations provides affordances internal representations lack.

Architects, designers, and engineers exploit the benefits of persistence and material affordance when they build models. Models have a special role in thinking, and can for our purposes be seen as two, three, or even four-dimensional external representations: paper sketches—2D; cardboard models, cartoons, and fly throughs—3D models in space or time; and dynamically changing three-dimensional spatial structures—4D models. To see the extra power these sort of external representations offer let us look at the scale models that architects build.

Scale models are tangible representations of an intended design. They serve several functions:

1. They can serve as a shared object of thought because they are logically and physically *independent* from their author. They can be manipulated, probed, and observed independently of their author's prior notion about how to interact with the model. This is vital for talking with clients, displaying behavior, functionality and detecting unanticipated side effects. It makes them public and intersubjective.
2. Models enforce consistency. The assumption behind model theory in mathematics is that if a physical structure can be found or constructed, the axioms that it instantiates must be consistent (Nagel and Newman 1958). Unlike a description of the world, or a mental representation, any actual physical model must be self-consistent. It cannot refer to properties that are not simultaneously realizable, because if it is a valid model it counts as an existence proof of consistency. In a many part system, part A cannot be inconsistent with part B if they both can simultaneously be present in the same superstructure. Similarly, the movement of part A cannot be inconsistent with the movement of part B if the two can be run simultaneously. Build it, run it, and thereby prove it is possible. Inconsistency is physically unrealizable. There are few more powerful ideas than this.
3. Models reveal unanticipated consequences. To say that an external model is independent of its creator is to emphasize that other people can approach the model in ways unconstrained by its creator's intention. Once a structure is in the public domain it has a life if its own. This is well appreciated in the case of ambiguous objects. Look at Fig. 10.7a. Its author may have intended it to be a convex cube with two concave sides extending to the bottom right. But a viewer may initially see those concave sides as a convex cube with a corner pointing outward. Look at the image longer and other interpretations should appear. Studies on mental imagery have shown that subjects who have not yet detected an ambiguity by the time they create a mental image are not likely to realize the ambiguity inherent in their image (Chambers and Reisberg 1985). It is as if they are sustaining their image under an interpretation, a prior conception. And so, they are closed to new interpretations. When externalized and in the visual field, however, the very processes of vision—the way the eye moves and checks for consistency—typically drives them to see the ambiguity.[1] When a structure is probed deeply enough, relations or interactions between parts, that were never anticipated may be easy to

[1] Although this argument concerns visual processing, it applies equally well to the physical interactions we can perform on the physical object.

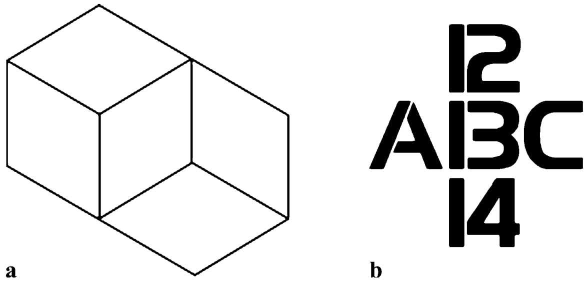


Fig. 10.7 Here are some ambiguous objects. In 10.7**a**, a variant of the Necker cube is shown where some corners that start looking convex (outward pointing) change to concave. 10.7**a** is ambiguous in several ways. Can you see at least four interpretations? In 10.7**b**, the middle element will seem to be a *B* or *13* depending on whether you read vertically or horizontally. How an object visually appears depends on how an agent looks at it, and this can be affected by how the structure is framed, how it is contextualized, how the agent feels, or what an agent is primed to see

discover. Thus, an author may be able to discover interpretations he or she never considered. Whether the thing externalized is a representation of a thought, image, or mental animation, its persistence and independence means that it may be reconsidered in a new light, and interacted with in a new manner.

The power of modeling is a topic of its own. Another special property is one that is made explicit in mathematical simulations that can be run back and forth under a user's control. Such simulations provide persistence and author independence because they can be run forward, slowed down, stopped, or compared snapshot by snapshot. All normal sized models support our physical interaction. We can move them in ways that exposes otherwise hard to see perspectives and relations. When our interaction is controlled precisely, or interpreted as movement along a timeline, we can juxtapose snapshots in time for comparisons that would simply be impossible otherwise. Without the stability of reproducibility and persistence, some of the ideas we form about the temporal dynamics of a structure would be virtually unthinkable.

Reformulation and Explicitness

A fourth source of the power of interaction relies on our ability to externally restate ideas. Sometimes it is easier to perform restatement externally than in our heads.

Representations encode information. Some forms encode their information more *explicitly* than others (Kirsh 1990). For example, the numerals '$\sqrt{2209}$' and '47' both refer to the number 47 but the numeral '47' is a more explicit encoding of 47. Much external activity can be interpreted as converting expressions into more explicit formulations, which in turn makes it easier to 'grasp' the content they encode. This is a major method for solving problems. For instance, the problem

$x = \sqrt[4]{28{,}561} + \sqrt{2209}$ is trivial to solve once the appropriate values for $\sqrt[4]{28{,}561}$ and $\sqrt{2209}$ have been substituted, as in $x = 13 + 47$.[2]

Much cognition can be understood as a type of external epistemic activity. If this seems to grant the theory of extended mind (Clark 2008) too much support add the word 'managing' as in 'much cognition involves managing external epistemic activity'. We reformulate and substitute representations in an effort to make content more explicit. We work on problems until their answer becomes apparent.

The activity of reformulating external representations until they encode content more transparently, more explicitly, is one of the more useful things we do outside our heads. But why bother? Why not do all the reformulation internally? A reason to compute outside the head is that outside there are public algorithms and special artifacts available for encoding and computing. The cost structure of computation is very different outside than inside. Try calculating $\sqrt{2209}$ in your head without relying on a calculator or an algorithm. Even savants who do this 'by just thinking' find there is a limit on size. Eventually, whoever you are, problems are too big or too hard to do in the head. External algorithms provide a mechanism for manipulating external symbols that makes the process manageable. Indeed were we to display the computational cost profiles (measured in terms of speed accuracy) for performing a calculation such as adding numbers in the head vs. using algorithms or tools in the world, it would be clear why most young people can no longer do much arithmetic in their heads. Tools reshape the cost structure of task performance, and people adapt by becoming dependent on those tools.

A second reason we compute outside rather than inside has to do with a different sort of complexity. One of the techniques of reformulation involves substitution and rewriting. For instance, if asked to find the values of x given that $x^2 + 6x = 7$, it is easiest if we substitute $(x + 3)^2 - 9$ for $x^2 + 6x$. This is a clever trick requiring insight. Someone had to notice that $(x + 3)^2 = x^2 + 6x + 9$, which is awfully close to $x^2 + 6x = 7$. By substituting we get $(x + 3)^2 = 16$, which yields $x = 1$ or -7. Could such substitutions be done in memory? Not likely. Again, there are probably some people who can do them. But again, there always comes a point, where the requisite substitutions are too complex to anticipate the outcome 'just by thinking' in one's head. The new expressions have to be plugged in externally, much like when we swap a new part for an old one in a car engine and then run the engine to see if everything works. Without actually testing things in the physical world it's too hard and error prone to predict downstream effects. Interactions and side effects are always possible. The same holds when the rules governing reformulation are based on rewrite rules. The revisions and interactions soon become too complex to expect anyone to detect or remember them.

[2]Reformulation is not limited to formal problem solving. The statement "Police policed police police police" is easier to understand when restated at "Police who are policed by police, also police other police". Most people would not break out their pens to make sense of that statement, but few of us can make sense of it without saying the sentence out loud several times.

a b c

Fig. 10.8 Imagine hearing 12 seconds of music. Now look at the musical notation shown here. Notation has the value of showing in space a structure that one hears. But there is much more in the sound as heard than is represented in the notation alone. Sound is the natural representation of music. The same is true for dance. Compare Laban notation for dance with the full body structure of dancers. Even if the joint structure is captured in Laban, how well represented are the dynamics of movement, the feel of the dance, and its aesthetic impression?

Natural Encoding

Persistence, reordering, and reformulation largely explain why externalizing information and representation may increase the efficiency, precision, complexity and depth of cognition. And if these aspects of interaction with external representations do not explain the extra power to be had then simulation does. Still, there is another aspect to consider: how external processes may increase the *breadth* of cognition. To explore this aspect consider again, why we prefer one modality to another for certain types of thinking.

Every representational system or modality has its strengths and weaknesses. An inference or attribute that is obvious in one system may be non-obvious in another. Consider Fig. 10.8—a musical notation. The referent of the notation is a piece of music. Music is sound with a specific pitch or harmony, volume, timber, and temporal dynamics. The 'home' domain of music, therefore, is sound. Visual notation for music is parasitic on the structure of sound. Prima facie, the best representation to make sense of musical structure is music itself; we go to the source to understand its structure.[3]

If there are times when the source medium is required to represent the content of a thought, then a further reason to externalize content and manipulate it outside is that for some problems, the natural representation of the content only exists outside. Arguably, no one—or at best only a few people—can hear music in their head the way it sounds outside. Mental images of sounds have different properties than actual sounds. Even if it is possible for the experience of the mental image of music to be as vivid and detailed as perception of the real thing, few people—other than the musically gifted, the professional musician, or composer (Sacks 2008)—can

[3]To see why music can be both referent and representation (terrain and also map) ask whether there is a difference between hearing sound and hearing sound *as* music. The sound is the terrain; music is the conceptualizing structure that interprets the sound; it maps it.

accurately *control* musical images in their heads. It is far easier to manifest music externally than it is to do so internally. So, for most people, to make sense of music the first thing to do is to play it or listen to it.

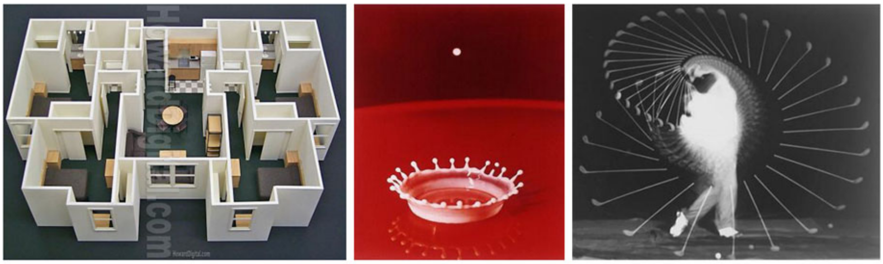
This raises a further requirement on the elements of thought. If a representational system is to function as a medium of thought the elements in the system must be sufficiently manipulable to be worked with quickly. Spoken and written words are malleable and fast. Body movements for dance, gesture, and perhaps the pliability of clay are too. Musical instruments, likewise, permit rapid production of sound. These outer media or tools for creating media support fast work. They enable us to work with plastic media. In this respect they enable us to work outside much like the way we work inside, using visual or auditory images for words or ideas, which most of us work with at the speed of thought. If external manipulability matches the internal requirements on speed, then an external medium has the plasticity to be a candidate for thinking in.

Using Multiple Representations

Despite the value of listening to music there are times when notation does reveal more than the music one has listened to—instances where a non-natural representation can be more revealing and intuitive than the original representation. Because a notational representation uses persistent, space consuming representations, early and later structures can be compared, superimposed and transformed using notation specific operators. As with logic and jigsaw puzzles it is useful to have tangible representatives that can be manipulated. In these cases, a subject who moves from one representation to the other may extend cognition. By moving between listening to music, and writing it down in a notation, or listening and then reading the notation, or sometimes vice versa, a composer or listener may be able to explore certain elements of musical structure that are otherwise inaccessible. The more complicated the structure of the music the more this seems to be true. Without interacting with *multiple representations* certain discoveries would simply be out of reach. Visual designers who move between pen and paper, 3D mockups and rapid prototypes are familiar with the same type of process.

Construction and Tools

The final virtue of external interaction I will discuss is, in some ways, the summation of persistence, rearrangement, and reformulation. It may be called the power of construction. In making a construction—whether it be the graphical layovers of the dancer shown in Fig. 10.5, the geometric construction of Fig. 10.1, or building a prototype of a design as in Fig. 10.9a—there is magic in actually making something in the world. As mentioned in the discussion of scale models, by constructing a structure we prove that its parts are mutually consistent. If we can build it, then it



a b c

Fig. 10.9 A 3D model permits architects to view a form from arbitrary angles. It allows them to measure, compare, and look for violation of constraints. By approaching the model from odd angles they can see occlusions, and relations that would be extremely hard to see otherwise. In 10.9**b** we see a near perfect coronet formed by a drop of milk in a famous photograph by Harold Edgerton. And in 10.9**c** we see a famous stop frame image of a golfer swinging and then hitting a golf ball (Densmore Shute Bends the Shaft, 1938, © Dr Harold Edgerton, Silver Gelatin Print)

must be logically and physically viable. If we can run it, then the actions of those parts are consistent, at least some of the time; and if we can run it under all orderings then it is consistent all of the time. The physical world does not lie.

The constructive process has a special place in human thinking because it is self-certifying. In mathematics, constructive reasoning means proving a mathematical object exists by showing it. For example, if it were claimed that a given set has a largest element, then a constructionist proof would provide a method for finding the largest element, and then apply the method to actually display the element.

Not every form of human reasoning is constructive. Humans reason by analogy, by induction, they offer explanations, and they think while they perform other activities, such as following instructions, interpreting a foreign language, and so on. None of these are constructive methods in the mathematical sense. However, because of the incremental nature of construction the effort to construct a solution may also be a way of exploring a problem. When students look for a constructive proof to a geometric problem, they use the evolving external structure to prompt ideas, bump into constraints, and realize possibilities. When they write down partial translations of a paragraph they rely on explicit fragments to help guide current translation.

The question that begs to be asked is whether thinking with external elements is ever necessary. Can we, in principle, do everything in our heads, or do we need to interact with something outside ourselves in order to probe and conceptualize, and get things right? In mathematics, externalization is necessary, not just for communication, but to display the mathematical object in question. It is like measurement: you cannot provide the value of a physical magnitude without measuring it. You cannot show the reality of a mathematical object (for constructivists) without revealing a proof that parades it. Yet, during the discovery process might not all the thinking be internal, the result of an interaction between elements inside the head? Where is the proof that, at first, all that probing and conceptualizing might not be the outcome of a purely internal activity? Even if the internal activity is simulating what it would be like to write things down outside, or how one would present one's idea to others, all the ‘real’ thinking lives internally. We needed the outside world to teach us how to think,[4] but once we know how, we never need to physically encounter tangible two or three-dimensional structures to epistemically probe the ‘world’.

I believe this is wrong: physical interaction with tangible elements is a necessary part of our thinking process because there are occasions when we must harness physical processes to formulate and transition between thoughts. There are cognitive things we can do outside our heads that we simply cannot do inside. On those occasions, external processes function as special cognitive artifacts[5] that we are incapable of simulating internally.

To defend this hypothesis is harder than it might seem. In practice, few people can multiply two four-digit numbers in their heads. And, if they can, then increase the problem to ten digit numbers. This in practical limitation does not prove the ‘in principle’ claim, however, that normal human brains lack the capacity to solve certain problems internally that they can solve with external help, with tools, computers or other people. There are chess masters who can play equally well, or nearly as well, blindfolded as open eyed (Chabris and Hearst 2003). There is no evidence that a team of chess players is better than an individual. There are people with savant syndrome who can multiply large numbers in their head, or determine primes or square roots. Other savants with eidetic memories can read books at a rate of 8–10 seconds per page, memorizing almost everything.[6] Tesla said that when he was designing a device, he would run a simulation of it in his head for a few weeks to see which parts were most subject to wear (Hegarty 2004, p. 281, citing Shepherd). Stephen Hawking is said to have developed analytical abilities that allowed him to manipulate equations in mind equivalent to more than a page of handwritten manipulations. For any reasoning problem of complexity n, how do we know there is not some person, somewhere, who can solve it in their head, or could, if trained long enough? To be sure, this says little about the average person. Any given person may reach their computational limit on problems much smaller than n. And our technology and culture has evolved to support the majority of people. So, in practice, all people rely on available tools, practices, and techniques for reasoning. Nonetheless, if a single person can cope with n, then there is an existence proof that the complexity of external simulation does not itself mean that internal simulation is

[4]Vygotsky among others has suggested that we mastered thinking externally, by conforming our behavior to social norms of rational inquiry, and that what we learned to do first on the outside we came to do on the inside. Thus, the reason we can do math in our head is because we can do math in the world. The same applies to thinking internally in auditory images. We think in words internally, using auditory images of sounds, because when we think in public we speak. Thinking internally is simulating what we do externally, though Vygotsky did believe that inner speech of adults would be much compressed and unintelligible to anyone except the thinker (Vygotsky 1986).

[5]Hutchins (2001).

[6]Entry from Wikipedia on Kim Peek (2009) the inspiration for the character in the movie *Rain Man*: “He reads a book in about an hour, and remembers almost everything he has read (...) His reading technique consists of reading the left page with his left eye and the right page with his right eye and in this way can read two pages at time with a rate of about 8–10 seconds per page. He can recall the content of some 12,000 books from memory.

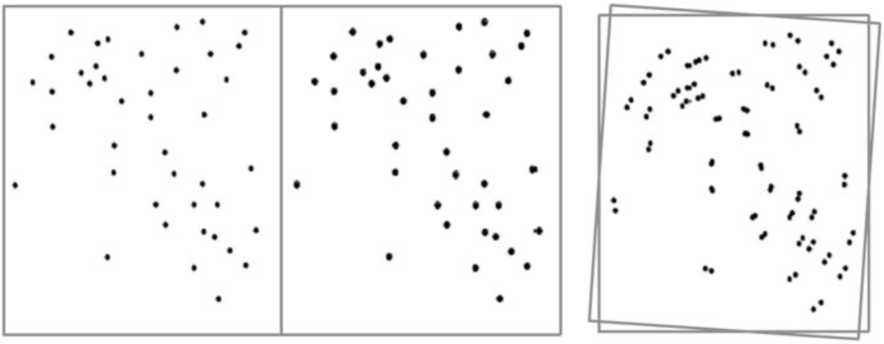


Fig. 10.10 On the *left* are two collections of *random dots*. They differ only by the rotation of the plane they are in. On the *right*, they have been superimposed. Their global relation is now visible. Could this relationship be detected without physically superimposing the patterns? Mental imagery does not support vivid superimposition. And if there are outlier humans who have this odd ability they will necessarily fail as the number of dots or the number of superimpositions increase

not possible. It suggests that any problem we cannot solve in our heads that we can solve with external help, has more to do with cost structure than with an in principle biological inability.

One way of making the in principle case is to show that there are *operations* that can be performed on external representations that cannot be performed on internal representations, and that, somehow, these are essential. Are there epistemic activities we can perform outside that we cannot duplicate inside, not because of their complexity, but because there are physical properties and technologies available on the outside that we cannot duplicate mentally—operations we cannot mentally simulate with sufficient realism to deliver dependable answers?

Consider Fig. 10.10. The dots in the two images on the left are related to one another by a rotation of 4°. This is essentially invisible unless the two images are superimposed, as in the image on the right. Superimposition is a physical relation that can be repeated any number of times, as is rotation. Both require control over physical transformations. In the case of superposition the position of the layers must be controlled precisely, and in the case of rotation, the angle must be controlled precisely. Are there such functions in the brain?

The process required in the brain is analog. For over 25 years, a dispute has raged over whether brains support analog processes or whether mental imagery is driven by non-analog means (Pylyshyn 2001). We can sidestep this question, though, by appealing to an in principle distinction between types of processes. In an important paper, Von Neumann (1948 [1961]) mentioned that some processes in nature might be irreducibly complex. Any description of one of those processes would be as complex as the process itself. Thus, to simulate or model that process one would have to recreate all the factors involved. This holds regardless of whether the simulation or modeling is being performed internally or externally. Von Neumann put it like this:

> "It is not at all certain that in this domain a real object might not constitute the simplest description of itself, that is, any attempt to describe it by the usual literary or formal-logical method may lead to something less manageable and more involved" (p. 311).

Marr (1977) invoking the same idea, spoke of Type 2 processes where any abstraction would be unreliable because the process being described evolves as the result of "the simultaneous action of a considerable number of processes, whose interaction is its own simplest description" (p. 38). Protein folding and unfolding are examples of such processes, according to Marr.[7] Other examples might be the *n* body problem, the solution to certain market equilibrium problems, situations where the outcome depends on the voting of *n* participants, and certain quantum computations.

The hallmark of these problems is that there exists physical processes that start and end in an interpretable state, but the way they get there is unpredictable; the factors mediating the start and end state are large in number, and on any individual run are impossible to predict. To determine the outcome, therefore, it is necessary to run the process, and it is best to run the process repeatedly. No tractable equation will work as well.

How are these problems to be solved if we have no access to the process or system itself? The next best thing is to run a physically similar process. For example, to compute the behavior of an *n* body system, such as our solar system, our best hope is to construct a small analog version of that system—an orrery—then run the model, and read off the result (see Fig. 10.11). Using this analog process, we can compute a function (to a reasonable degree of approximation) that we have no other reliable way of computing.

The implication is that for brains to solve these sort of problems, it would be necessary for them to encode the initial state of the Type II system, and then simulate the physical interaction of its parts. If this interaction is essentially physical—if for instance, it relies on physical equilibria, or mechanical compliance, or friction—there may be no reliable way of running an internal simulation. We need the cognitive amplification that exploiting physical models provides. We would need to rely on the parallel processing, the physical interaction, and the intrinsic unpredictability of those analog systems.

The conclusion I draw is that to formulate certain thoughts and to transition to others, we must either be able to represent arbitrarily complex states—states that

[7]From Marr (1977, p. 38) "One promising candidate for a Type 2 theory is the problem of predicting how a protein will fold. A large number of influences act on a large polypeptide chain as it flaps and flails in a medium. At each moment only a few of the possible interactions will be important, but the importance of those few is decisive. Attempts to construct a simplified theory must ignore some interactions; but if most interactions are crucial at some stage during the folding, a simplified theory will prove inadequate. Interestingly, the most promising studies of protein folding are currently those that take a brute force approach, setting up a rather detailed model of the amino acids, the geometry associated with their sequence, hydrophobic interactions with the circumambient fluid, random thermal perturbations etc., and letting the whole set of processes run until a stable configuration is achieved (Levitt and Warshel 1975)."

Fig. 10.11 In this mechanical orrery by Gilkerson, housed in the Armagh Observatory the movement of the planets and their moons are mechanically simulated. It is not possible to access an arbitrary position of the system without moving through intermediate states. This is a feature of simulation systems: they do not have a closed form or analytic solution. To compute the state of the system at t_{12} one must determine the state at t_{11} and move from there

cannot be represented in compact form—or we must rely on the external states themselves to encode their values and then use them to transition to later states. These external states we are able to name but never characterize in full structural detail.[8]

Conclusion

In order to extract meaning, draw conclusions, and deepen our understanding of representations and the world more generally, we often mark, annotate and create representations; we rearrange them, build on them, recast them; we compare them, and perform sundry other manipulations. Why bother? Minds are powerful devices for projecting structure on the world and imagining structure when it is not present. Our inner mental life is plastic and controllable, filled with images of speech, visual scene, and imageless propositions. For most of intellectual history this impressive capacity has been assumed sufficient for thought. Why do we bother to interact so much?

I have argued that much of thinking centers on interacting with external representations, and that sometimes these interactions are irreducible to processes that can be simulated, created, and controlled in the head. Often, the reason we interact with external representations, though, boils down to cost. Nothing comes without a cost.

[8]In practice, though not in principle, computers fall into this category. When a workplace has been augmented with tools such as wizards, software agents and the like, it is possible to multiply the potency of basic strategies of interaction to the point where such increases qualitatively change what humans can do, what they can make sense of, and so on. Sometimes our best tools are analog, however, and these are the ones that may provide in principle augmentations to human thought.

A useful approach to understanding epistemic interaction is to see it as a means of reducing the cost of projecting structure onto the world. To solve a geometric problem we might imagine a structure and reason about it internally, we might work with an illustration and project extensions and possibilities. At some point though the cost of projection becomes prohibitive. By creating external structure that anchors and visually encodes our projections, we can push further, compute more efficiently, and create forms that allow us to share thought. I have presented a few of the powerful consequences of interaction. It is part of a more general strategy that humans have evolved to project and materialize meaningful structure.

References

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.

Chabris, C. F., & Hearst, E. S. (2003). Visualization, pattern recognition, and forward search: effects of playing speed and sight of the position on grandmaster chess errors. *Cognitive Science*, *27*, 637–648.

Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology. Human Perception and Performance*, *11*, 317–328.

Clark, A. (2008). *Supersizing the mind: embodiment, action, and cognitive extension*. Oxford: Oxford University Press.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, *8*, 280–285.

Hutchins, E. (2001). Cognitive artifacts. In R. Wilson & F. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences*. Cambridge: MIT Press.

Hutchins, E. (2005). Material anchors for conceptual blends. *Journal of Pragmatics*, *37*, 1555–1577.

Forsythe, W. (2008). http://www.youtube.com/watch?v=0P_4D8c2oGs&feature=related.

Kirsh, D. (1990). When is information explicitly represented? In *The Vancouver studies in cognitive science* (pp. 340–365). Oxford: Oxford University Press.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, *73*, 31–68.

Kirsh, D. (1996). Adapting the environment instead of oneself. *Adaptive Behavior*, *4*, 415–452.

Kirsh, D. (2009a). Problem solving and situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 264–306). Cambridge: Cambridge University Press.

Kirsh, D. (2009b). Projection, problem space and anchoring. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2310–2315). Austin: Cognitive Science Society.

Kirsh, D. (2009c). Interaction, external representation and sensemaking. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 1103–1108). Austin: Cognitive Science Society.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic actions. *Cognitive Science*, *18*, 513–549.

Levitt, M., & Warshel, A. (1975). Computer simulation of protein folding. *Nature*, *253*, 694–698.

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove: Erlbaum.

Marr, D. (1977). Artificial intelligence: a personal view. *Artificial Intelligence*, *9*, 37–48.

Nagel, E., & Newman, J. (1958). *Gödel's proof*. New York: New York University Press.

Pylyshyn, Z. (2001). Is the imagery debate over? If so, what was it about? In E. Dupoux (Ed.), *Language, brain, and cognitive development: essays in honor of Jacques Mehler* (pp 59–83). Cambridge: MIT Press.

Quine, W. V. O. (1960). *Word and object*. Cambridge: MIT Press.
Sacks, O. (2008). *Musicophilia: tales of music and the brain*. New York: Knopf.
Strawson, P. F. (1959). *Individuals: an essay in descriptive metaphysics*. London: Methuen.
Von Neumann, J. (1948 [1961]). General and logical theory of automata. In A. H. Taub (Ed.), *John von Neumann: collected works. Vol. V: Design of computers, theory of automata and numerical analysis* (pp. 288–326). Oxford: Pergamon Press.
Vygotsky, L. (1986) *Thought and language* (rev ed.). Cambridge: MIT Press.
Wikipedia on Kim Peek (2009). http://en.wikipedia.org/wiki/Kim_Peek.
Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell Sci.

Chapter 11
Human Interactivity: Problem-Solving, Solution-Probing and Verbal Patterns in the Wild

Sune Vork Steffensen

Abstract This chapter presents an interactivity-based approach to human problem-solving in the wild. It introduces the notion of 'interactivity', here defined as sense-saturated coordination that contributes to human action. On this view, interactivity is an ontological substrate that can be studied as interaction, as cognition, or as ecological niche production. While the chapter theoretically argues in favour of a unified, trans-disciplinary approach to interactivity, it turns its attention to the cognitive ecology of human problem-solving. It does so by presenting a method of Cognitive Event Analysis, that leads to a detailed analysis of how a problem in the wild is being solved. The analysis addresses the cognitive dynamics of how two persons in a work setting reach an insight into the nature of a problem. These dynamics include the spatial organisation of the workplace, the interbodily dynamics between the two participants (especially in relation to gaze and the manual handling of papers), and verbal patterns that prompt them to simulate how the problem appears to a third party. The chapter concludes that human problem-solving is far less linear and planned than assumed in much work on the topic. Rather that problem-solving, it appears as solution-probing in real-time. The cognitive trajectory to a viable solution is thus self-organised, unplanned, and on the edge of chaos.

Introduction

The biggest problem in finding a needle in a haystack is that in most haystacks there are no needles. Comparably, in problem-solving the real challenge is problem-finding, that is identifying the nature of the problem. In problem-finding, our everyday experience differs immensely from what is being examined in the laboratory settings that nurture cognitive psychology. In this artificial setting subjects are typically presented to a well-defined problem; for example "connect the nine dots with four connected straight lines without lifting your pencil from the paper" (Weisberg and Alba 1981, p. 170; the nine dot problem was originally presented in Maier

S.V. Steffensen (✉)
Centre for Human Interactivity, Institute of Language and Communication, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
e-mail: s.v.steffensen@sdu.dk

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_11,

1930). In contrast, real-life existence does not provide us with the luxury of pre-established, well-defined problems. We rarely meet instructions like "go down to the diner on the corner, order a cheese burger, bring it back to your office and eat it." We just feel hunger. From that feeling, we need to figure out what to do. Human life unfolds in an indefinite problem space. Hutchins's (1995a, 1995b) statement is correct: *Cognition in the wild* is very different from the laboratory activities that cognitive psychologists for a century have investigated as problem-solving. However, as this chapter demonstrates, in most real-life contexts, *cognition in the wild* is even wilder than generally assumed in the Distributed Cognition literature. Rather, it is profoundly interpenetrated with the real-time flow of human co-existence in a distinct human ecology as human beings engage in collective and socioculturally enhanced problem-searching, problem-finding and problem-solving in a self-organised problem space. Following Kirsh (1997), Steffensen (2012), Pedersen (2012), and Vallée-Tourangeau et al. (2011), I refer to this real-time flow of co-existence as *interactivity*. While the concept is clarified below, interactivity is defined as *sense-saturated coordination that contributes to human action*.

As interactivity comprises all aspects of this flow of co-existence, it has a primordial quality that does not fit readily into how science is compartmentalised. Thus, nowhere has the luxury of predefined problem spaces been more emphasised than where disciplines each seek to grasp their small part of reality.[1] Thus, what the cognitivist perceives as problem-solving, the microsociologist interprets as social interaction, and the biologist construes as human ecological niche construction. Underlying each perspective, one finds interactivity: it is an ontological substrate that each discipline has turned into an 'object'. While all three perspectives may yield descriptively adequate models within an epistemological domain (cf. Hemmingsen 2013, Chap. 6), they cannot, in themselves, provide an explanatory model of interactivity, of what really happens in the flow of human existence. Though we should not delude ourselves by believing that we are even close to such a model, a focus on interactivity may provide a grounding for a theoretical and methodological approach to what Andy Clark has called *The Heideggerian Theater*: "Our target is not just a neural control system but a complex cognitive economy spanning brain, body and world. Within this complex economy, the body plays a crucial role. [...] The body is—dare I say?—the Heideggerian Theater: the place where it all comes together, or as together as it comes at all" (Clark 2008, p. 217).

In contrast to Clark, and many recent approaches in cognitive science, a focus on interactivity rejects organism-centrism. It does not pivot on the living *body* but, rather, on *bodies-in-action*. Accordingly, rather than appeal to the utility functions of a cognitive *economy*, the emphasis turns to how we act viably in a cognitive *ecology*. I thus maintain that this approach is better equipped for actually responding to Clark's desideratum or the aspiration to rethink human nature as "not accidentally

[1] Saussure was indeed correct when he, a century ago, established that "c'est le point de vue qui crée l'objet" (de Saussure 1916/1972, p. 23).

but profoundly and continuously informed by our existence as physically embodied, and as socially and technologically embedded, organisms" (Clark 2008, p. 217).[2]

This chapter begins with a theoretical discussion (in the section *Human Interactivity*) of interactivity, follows up with a methodological intermezzo (in the section *From Interactivity to Cognitive Events*) and then presents a case study of human interactivity (in the section *The Invoice Case*). Finally, I draw some conclusions on how the approach can contribute to understanding of language and cognition.

Human Interactivity

Defining interactivity as sense-saturated coordination that contributes to human action characterises three aspects of the relevant phenomena. First, *coordination* refers to a reciprocal flow of minuscule, pico-scale interbodily movements that link and lock human beings in self-organised systems. This basic insight is shared by work that invokes concepts such as *distributed cognitive systems* (Hollan et al. 2000), *situated activity systems* (Goodwin 2000), and *dialogical systems* (Steffensen 2012). Second, this coordination is *sense-saturated*, that is, it is pervaded by our species-specific capability for sense-making (Linell 2009). We engage in sense-making as our bodies integrate present circumstances with autobiographic memories and sociocultural histories: through sense-making the not-here and the not-now saturate our here-and-now coordination. Third, sense-saturated coordination constrains what we do and how we do it. For instance, if you grew up with the habit of greeting through cheek kissing, you know on which cheek to start, and you know how many kisses to exchange. But if you are not accustomed to the habit, and you greet a person who is, you can still engage in the social practice of greeting-through-cheek-kissing, but you are to a wide extent dependent on following the dynamics of the other person's body. This example illustrates the power of the interbodily dynamics, our interbodily agility as we mutually interpret and anticipate each other's movements, and the constraining and enabling function of socioculture. Because interactivity is sense-saturated, our actions and experiences are, at once, situated and non-situated, and we are furthermore bound to overthrow the monolithic distinction between the realm of sense-making (e.g., 'language') on the one side and the realm of behaviour on the other (cf. Steffensen 2011). Thus, what we normally conceive of as 'human language' is a pattern in interactivity, and as such it is always dynamic, as well as symbolic (Cowley 2011).

The interplay between interbodily dynamics and sociocultural constraints on behaviour prompts the approach to adopt a systemic focus on *results*, "not as a simple effect or consequence of behaviour but [as] a new possibility of action forming in the process of transition from one act to another" (Järvilehto 2009, p. 116). In focusing on results, the Heraclitean flow of interactivity ceases to be either a sequential

[2]For a fuller critique of Clark's position, see Steffensen (2009, 2011).

string of separate stimuli and responses or an incessant flow of undifferentiated, unstructured "pure being."

The concept of 'human interactivity' grounds trans-disciplinary empirical work that attempts to move beyond the microsocial study of social interaction, beyond the cognitive study of functional, computational systems and minds, and beyond the study of biological organisms *sui generis*. It presupposes a crucial distinction between 'interaction' and 'interactivity'; whereas interaction captures a relation of dependence between separable systems, interactivity explores their bidirectional coupling. This contrasts with a focus that seeks out the natural laws that describe, for example, how the earth and the moon interact (cf. Kirsh 1997, p. 83). But it also contrasts with approaches that invoke social rules in order to explain how separate human agents orient to each other and the world (cf. Hodges and Baron 1992). On this views, as social rules are violable, human interaction becomes purely normative. In ethnomethodology and conversation analysis, therefore, human interaction is a *contingent* co-construction of negotiated meaning. Interestingly, both celestial interaction (i.e., natural laws) and social interaction (i.e., social rules) can be studied by strict application of inductive methods. In both domains, theorists provide formulations or *models* of how inert bodies or social actors orient to each other (cf. Sacks et al. 1974). And just as astronomers need not claim any larger purpose in nature, interaction analysts avoid issues of intentionality. They neither allow biological bodies to *become* social actors or ask *why* people act as they do. Any method based on studying interactional regularities and repeated patterns will fail to capture the biological dynamics of human interaction or its (cognitive) *results*. They simply overlook variable aspects of the interactivity that connects living beings.

Interactivity takes us beyond the computational functionalism of cognitive science (including that of classical Distributed Cognition). Functionalists propose that computational processes can take place both inside and outside living beings, irrespective of the medium in which they unfold. By contrast, the concept of 'interactivity' views the "glue of cognition" (Kirsh 2006, p. 250) as "complex, dynamic coupling between two or more intelligent parties" (Kirsh 1997, p. 83). It thus makes a difference how a cognitive system balances animate agency with non-animate contributions. Thus, if the cognitive work is distributed between animate parties (human beings, to simplify the picture), "the involved parties must *co-ordinate* their activity or else the process collapses into chaos; all parties exercise *power* over each other, influencing what the other will do, and usually there is some degree of (tacit) *negotiation* over who will do what, when and how" (Kirsh 1997, pp. 82–83; emphasis in the original). Cognition depends on the total organisation of organisms' self-reflexive being in their shared environment. On this view, human *interactivity* is a non-local phenomenon (Steffensen and Cowley 2010): it links the biological organism with autobiographical memory, sociocultural resources (e.g., verbal patterns) and environmental structures. Interactivity is thus whole-bodied activity that flows between human beings—or between a single human being and cultural artefacts and procedures (e.g., previously crafted texts or technological devices). To generate *joint results*, it is enough that we coordinate and strive to make things happen.[3]

Along with other contemporary work in cognitive science, this approach challenges biological reductionism. As living organisms perceive *as* they act and, conversely, act *as* they perceive; far from being an inner process, cognition exploits the full organisation of embodied and emotional action-perception cycles (Kirsh 1997; Järvilehto 1998; Robbins and Aydede 2009). This view thus traces cognition to a history of how living systems adapt within a changing environment. Situated, living, cognitive systems thus function across the boundary of the skin; accordingly, we take an *ecological* stance that overthrows the dichotomy of computational reductionism and biological reductionism. While the former reduces cognition to its functional properties, overlooking feelings and the properties of bodies, the latter stick to dermal metaphysics: on a priori grounds, the skin is taken to define the boundary of what counts as cognitive. In contrast, a systemic approach to interactivity pivots on an extended ecology (Steffensen 2011) that defines human cognition in terms of the flexibility and adaptivity of (human) organism-environment-systems (Järvilehto 1998, 2009).

A focus on interactivity obliterates any sharp boundary between the biological and the social. First, social normativity informs and constrains our biological being, for example by filtering out activities that are judged inappropriate in certain social contexts (e.g., picking one's nose). Second, and crucially, our very existence depends on conspecifics and the extended ecology. Human development depends on caregivers (e.g., Trevarthen 1998; Bråten 2009) who enable the new-born infant to learn from being tightly coupled to human beings in its environment. Plainly it is unwarranted to limit biology to the body: as people breathe, move, touch, act, perceive and care, they *radiate* into their surroundings. Our being is interbodily being.

From Interactivity to Cognitive Events

Interactivity is a primordial substrate of human life. Phenomenologically, we may perceive it as 'language', 'interaction', 'cognition' or 'niche-construction', but these are only *perspectives* on interactivity, not ontologically real phenomena *per se*. Having made this claim, I now turn to how it serves human beings as they find and solve problems in the wild, that is, in the real-time flow of day-to-day existence. Accordingly I focus on the *cognitive* dynamics of interactivity that take part in what Hollan et al. (2000) call *Distributed Cognitive Systems* (henceforth DCS). The DCS is a self-organising entity that arises as human beings co-engage through interactivity,

[3] In line with this view, Donald (2001) argues that humans alone developed a cultural capacity to *voluntarily* retrieve experience. Culture turns our social and physical environment into a *cognitive* resource, and not just an independent, external material resource. Therefore, studying cognition as a DCS requires that we take into account longer time-scales than that of here-and-now situated interactions, for instance those involved in autobiographical memory (Donald 1991, 2001, 2012; cf. Cowley 2012), collective memory (Wertsch 2002) and cultural artefacts.

and connect up brains, bodies and aspects of the environment. The dynamics of the DCS derive from how one or more people engage with each other and external artefacts. Such a distributed system has emergent cognitive properties that differ radically from those of the components; moreover, these properties cannot be inferred from the properties of the components, no matter how much we know about the details of the properties of those components (cf. Hutchins 1995a, 1995b).

The cognitive emphasis also reflects in the term for the method applied in this chapter, namely *Cognitive Event Analysis* which builds on Steffensen et al. (2010), Steffensen (2012), Galosia et al. (2010), Pedersen (2010, 2012). The method of analysis proceeds in two stages: first, since interactivity lacks inherent beginnings or endings, empirical happenings are defined by external criteria. The first stage is therefore an *event identification* stage. For instance, in a medical context the presentation of a given patient symptom can function as such a criterion, irrespective of how the involved participants orient to it, and irrespective of whether it is noticed or not. The second stage tracks the dynamics involved in the interactivity by means of an *event trajectory analysis*. The two stages will be elaborated below and a hierarchy of DCS components is suggested.

Event Identification

Given its primordial status, the current approach defines interactivity neither in terms of what people do, say, or mean. Rather, it asks: *what happens?* Galosia et al. (2010) uses this question to define events, that is "*changes in the layout of affordances*" (Chemero 2000, p. 39; emphasis in the original) which yield *results*. This approach is indebted to Järvilehto's *systemic psychology* (or *organism-environment systems theory*), according to which "the research should start from the determination of the results of behaviour and lead to the necessary constituents of the living system determining the achievement of these results" (Järvilehto 2009, p. 118). Results are typically, though not necessarily, triggered by how people draw on sense-saturated coordination; they are, however, not necessarily related to a participant's desired outcomes.

Given a set of criteria that are relevant to the events (e.g., diagnosing a patient), the results are *identifiable*. However, there is no reason to think that interactivity depends on a determinate set of results—what happens will not be the same for different participants and investigators. By tying results to external criteria, they become a probe that serves to explore the cognitive ecology of human interactivity. The focus on results obviously links the approach to cognitive science, and indeed traditional and, by extension, lab-based cognitive psychology plays an important part when it comes to formulating external criteria that define results. From an experimental perspective, problem-solving can be seen as 'search' in a temporally organised problem space, where the lab subjects are requested to move from a 'before' to an 'after'. In real life, as illustrated below, this temporal dimension often appears *post hoc*. In other words, identifying temporal patterns that resemble a journey through a problem space give the event a "pseudo-task structure." A results-based approach to interactivity thus offers criteria for zooming in on interactivity to clarify how salient events unfold at moments of an unending process. In other words, this approach provides a method for identifying *how* cognitive events arise from the flow of human interactivity.

Event Trajectory Analysis

An *event trajectory* is the path taken by a DCS as it moves through an infinite problem space in a way that yields results. Assuming that distributed cognitive systems are *animated* by persons, such an analysis focuses on the dynamics of what people *do*, as they perform various tasks. Thus a cognitive event analysis clarifies how the DCS reconfigures to yield results and how these are generated. Reconfigurations and results act as *transition points* on the event trajectory: the pattern and timing of these transition points thus give each event trajectory a 'fingerprint'. In events where problems are solved, the emergence of the solution defines a transition point that divides the event trajectory into a 'before' (big problem, no solution) and 'after' (good solution, no problem). Such a transition point is more salient than other transition points: it constitutes the *event pivot* or cognitive origo. Its importance lies in how it alters the event trajectory: *before* the event pivot, the participants seek a solution to the problem; *after* the event pivot they react to what they have found. As the event pivot constitutes a cognitive origo, Cognitive Event Analysis uses the convention of timing the event trajectory in relation to the event pivot and in milliseconds (ms). Other important transition points in the event trajectory include (a) the first observation of the problem; and (b) how unsuccessful strategies are gradually replaced by more successful ones. The case study below gives a more thorough introduction to the event trajectory, event transition points, and the event pivot.

Components in Distributed Cognitive System

Since interactivity flows across human agents and non-human artefacts, it becomes vital to establish what does and does not, for a given event, function as parts of the cognitive system. Following Clark's (2008) *Principle of Ecological Assembly*, in which it is assumed that "the canny cognizer tends to recruit, on the spot, whatever mix of problem-solving resources [that] will yield an acceptable result with a minimum of effort" (p. 13), I define the reconfiguration of a DCS in terms of the inclusion and exclusion neural, bodily, worldly, virtual or historical structures. This description of the DCS as a plastic self-reconfiguring system resembles Wilson and Clark's (2009, p. 65) description of a "transient extended cognitive system," which they define as "a soft-assembled whole that meshes the problem-solving contributions of the human brain and central nervous system with those of the (rest of the) body and various elements of local cognitive scaffolding." The main differences between the position advocated here and that of Wilson and Clark, is, first, that their focus is on how the individual cognizer ("the human brain" in the singular) recruits bodily and environmental structures in the cognitive meshwork, while I focus on how the interpersonal dynamics in a dialogical system constrain the cognitive trajectory of the whole system (cf. Steffensen 2011). Second, the ecological-observational study presented here explores *how* such a system manages its self-reconfiguration, while Wilson and Clark seem to be more interested in the contributions of various (biological and non-biological) parts to the extension of the human mind. This, however, does not clarify *how* the plasticity of cognitive systems is achieved.

In spite of these differences, both positions seem to agree that there is no *a priori* basis for specifying what is part of—or can become part of—a cognitive system. Thus, when a DCS is said to consist of X, Y and Z structures, it is the result of a *post hoc* procedure where the event trajectory is translated into a static cognitive inventory. The circularity of this procedure is non-trivial: while an event trajectory is non-linear and unpredictable, the component structures that are (or can become) part of the DCS constitute affordances that *constrain* the infinite problem space of the DCS. This, indeed, is why the methodology presented here uses a heuristic in seeking out how solutions are reached. This heuristic consists of a tentative hierarchy of component structures that are likely to contribute to the DCS. First of all, as agreed on by both the TECS and the DCS approach, a cognitive system is animated by human beings, and therefore it requires human components (for at least some of its trajectory). Accordingly, persons are regarded as an especially salient part of a DCS. However, within that category, many heterarchical relations can be involved; for instance, one person may be organisationally responsible for directing the DCS, or two participants can take turns as main cognizer (cf. Galosia et al. 2010). Second, especially in the "highly rationalised environments" (McGarry 2005, p. 187) described in the Distributed Cognition literature, *technologies* that were specifically designed to serve in a DCS are likely to play a major role (at some stage of the event trajectory, at least). This applies, for example, in Hutchins's (1995b, p. 274) description of the airspeed indicator in a McDonnell Douglas MD-80. However, this cannot be generalised as a rule: an alarm system will do no good in the DCS if a human component has turned it off or fails to pay attention. Third, many kinds of *artefacts* are redeployed in the DCS, either because of their material affordances (cf. the painter who uses his brush to measure a scenery), or because human beings interpret them as giving off information. Perry (2013, Chap. 9) provides a good example: "if someone had a pair of muddy boots under their desk, it meant that they have been on the [construction] site and could be asked about the current work situation" (Perry 2013). Fourth, I stress Hollan et al.'s (2000, p. 176) insight that cognition "may be distributed through time in such a way that the products of earlier events can transform the nature of later events." In workplace settings, such temporally distributed parts of the DCS often appear as *procedures* that, like second order language (cf. Thibault 2011), constrain human behaviour. Procedures aim at standardising the performance, typically through written instructions, but these may also be oral or memorised. The inclusion of memorised procedures in this hierarchy

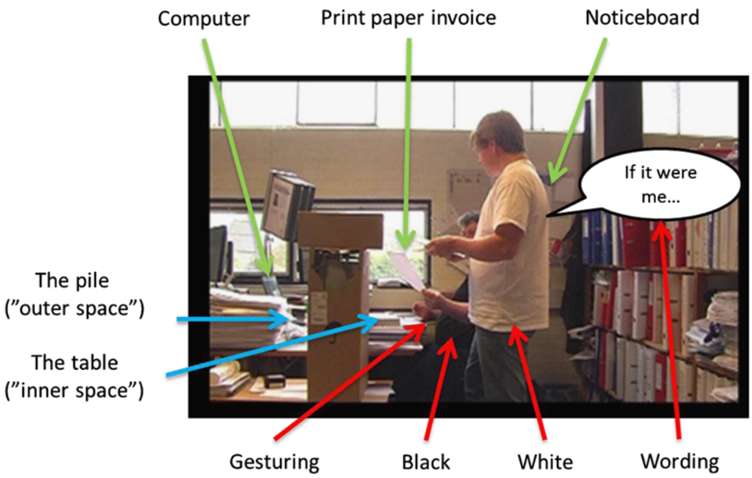


Fig. 11.1 The distributed cognitive system in the invoice case. The *four bottom* (*red*) *arrows* point to the human components of the DCS; the *three top* (*green*) *arrows* to the artefactual components; and the *two left* (*blue*) *arrows* to the spatial organisation of the DCS's environment

rests on an invert version of Clark's parity principle (Clark and Chambers 1998): if a standardisation counts as a procedure in its written form, we also accept memorised aids that afford uniformity in performance as procedure. Fifth, we can also include *narratives*. Narratives provide folk-theoretical schemes for understanding causality, ethics, emotions, for example. As such, narratives provide short term and long term constraints on behaviour.

The Invoice Case

The following data were collected by Anne-Kathrine Brorsen during her investigation of employer-employee interaction in a Danish company (Brorsen 2010). The data consist of 12 hours of video recordings and, in most of the data, the video recorder serves as a "silent observer" (Harel 1991)—it is placed on a tripod and records the same scenery for several hours. The scenery here is the workplace of one person sitting on a chair at his desk with a computer in front of him. From time to time, a co-worker appears and stands next to the first worker, while they interact about various issues. Both men are dressed in jeans and t-shirts, and I refer to them by the colours of their shirts. Thus, the man sitting at his work-station is *Black*, while his co-worker is *White* (see Fig. 11.1 for an overview of the lay-out).

Cognitive event analysis was used to identify a single instance of problem solving from the massive data set. It is thus a true instance of problem-solving in the wild: the two parties, led by Black, try to ensure that the company's invoices have the required Company Identification Number or '*CVR number*' (a CVR number is a code that the state agency *Det Centrale Virksomhedsregister* (*CVR*, The Danish

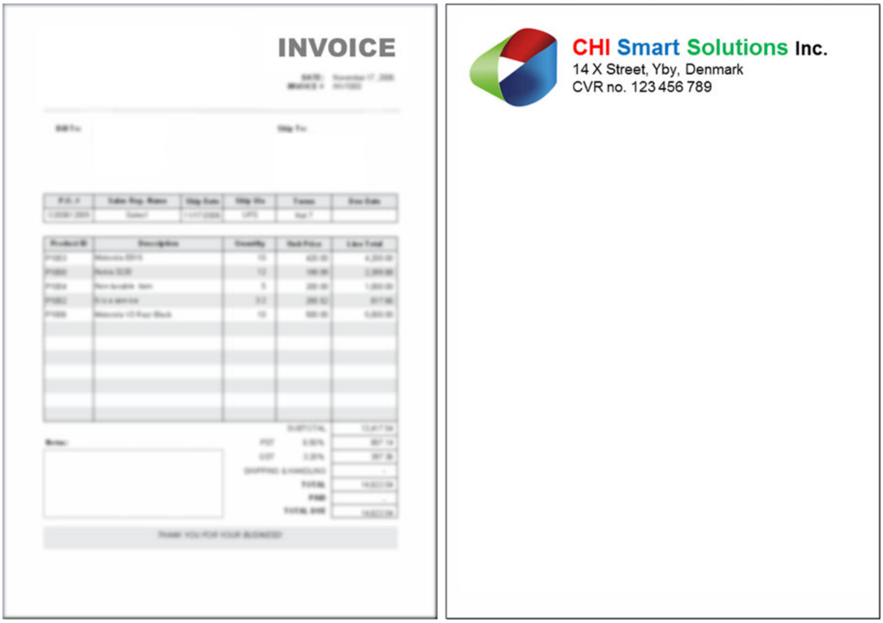


Fig. 11.2 The cognitive problem and its solution. To the *left* the figure shows a blurred reconstruction of the print paper invoice that the two men are handling. To the *right* it shows a reconstruction of the company's logo paper with the CVR number under the company's address

Central Business Register') supplies to all Danish companies). If this number does not appear on an invoice, it cannot be paid by the clients invoiced. In this situation, having collected a document from printer, no CVR number appears. This is the *problem* that Black and White faces as illustrated in Fig. 11.2. The left half of the figure reconstructs the printed invoice that the two men see; no CVR number appears. The right half is an anonymised reconstruction of the company's logo paper. As shown, the CVR number is in fact pre-printed under the company's name and address. Since the company would always use logo-headed paper when sending an invoice to a third party, there should be *no problem*. However, in preparing the invoice Black and While use what is printed on blank paper. Unbeknown to them, the problem arises from the choice of tray in the office printer! In principle, the *solution* is simple: they have to realise that they should discard the draft invoice for the version that their customers would receive. But as we all know from everyday life, problems are only simple when one knows the solution.

In this case, the salient cognitive event is identifying the problem: the event pivot arises in a 1900 ms sequence where White clearly articulates the solution: *nå nej men det er der jo hvis vi printer ud på logo papir* (Eng. 'well no but it [the number] is there if we print on logo paper'). According to the time notation the event pivot starts at 0 and ends at +1900. A second relevant event trajectory transition point arises as the participants shift from being unable to solve the problem, to reframing it in a way

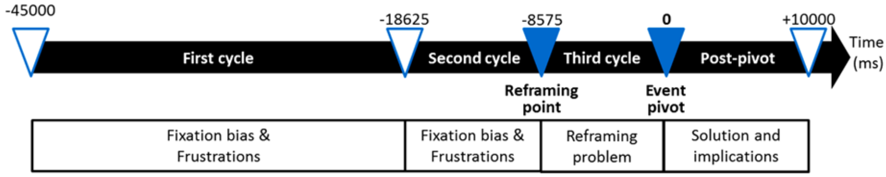


Fig. 11.3 The event trajectory. The figure shows the timeline of the event trajectory with indications of cycles and phases; *bold* (*blue*) *triangles* mark the two main transition points, the problem reframing point and the event pivot; *outlined* (*white*) *triangles* mark secondary transition points. The *boxes* indicate the main character of the phase/cycle in question

that gives a solution. This transition point occurs at −8575. Given these transition points, the event trajectory runs through the four phases shown in Fig. 11.3.

The first two phases (i.e. first and second cycle) are characterised by unsuccessful attempts to reach a solution. This ends at the problem reframing point. The third phase (or cycle) occurs between the transition points as the problem is reframed; finally, in the last phase, a solution emerges after the event pivot. Below I analyse the event trajectory in these four phases while giving emphasis to the cognitively salient problem reframing that occurs between the problem reframing point and the event pivots. The purpose of the analysis is to demonstrate that problem-finding, problem-solving and cognitive events do not just appear in an *aha*-like fashion: they depend on the flow of interactivity, as persons engage in solution-probing, i.e. they immethodically zigzag along a cognitive trajectory in an indefinite problem space, in their search for something that can work as a problem solution.

Frustration, Fixation and the Elusive Problem

As we enter the situation at −45000, Black and White are blocked by an ill-defined problem that seems to defy solution. Black repetitively restates *what* the perceived problem is. The pattern recurs three times, giving rise to three problem-solving cycles in the pre-pivot phases shown in Fig. 11.3. The first two cycles run from −45000 to −18625 (ca. 26 seconds) and from −18625 to −8575 (ca. 10 seconds), respectively, while the third run links the two transitions points, from −8575 to 0 (ca. 8.5 seconds). As in Steffensen et al. (2010), the presentation of these cycles takes a starting point in the micro-scale of the verbal patterns:

```
1. B: men (.) jeg kan fortælle dig der er ikke nogen som helst der
2.    vil betale den faktura der
3. W: nej det er jeg da godt klar over. [...]

1. B: but (.) I can tell you there is no-one whatsoever who will
2.    pay that invoice there
3. W: no I am aware of that. [...]
```

As shown by White's response in (3), he is already aware of the nature of the problem. At this moment, the information bearing properties of language—its symbolic function—is negligible. Often, it is simply false that language acts as "one of the structured representations produced and coordinated in the performance of the task" (Hutchins 1995b, p. 231). At best, it functions as a means of concerting their attention; at worst it takes them round in circles, preventing them from identifying the problem. Failure to come up with a solution frustrates the two participants as shown by two manifest outcomes. First, White acts out his frustration by blaming a party (unknown to the observer) that is referred to as *dem* (English: 'them').

```
3. W:                              [...]. Det er derfor vi sagde det
4.    til dem at den ikke duede jo. (0.7) Men det skulle vi ikke
5.    blande os i fordi det var som det var aftalt. (1.3)
6.    så var det det.
7. B: står der et CVR nummer hernede et sted?
8. W: =men det kan være du kan få et bedre svar end jeg ku,
9.    jeg fik bare at vide at sådan er det.
10.   [...]
```

```
3. W:                              [...]. That's why we told them
4.    that it was no good. (0.7) But that was not our business
5.    because it was as agreed upon. (1.3)
6.    So that was it.
7. B: is there a CVR number down here somewhere?
8. W: =but it might be that you can get a better answer than I
9.    could, I was just told that that was how it was
10.   [...]
```

Even weak versions of the so-called Frustration-Aggression Hypothesis (Dollard et al. 1939; Berkowitz 2007) would see this as predictable. Frustrations, it is found, tend to instigate aggressive or reproachful feelings by attributing the emotion to factors that lie beyond individual control. Second, and productively, the frustration instigates attempts at reframing. Thus, as White pauses in line 4, Black turns to his noticeboard on the right; then, 5400 ms after turning, he explains what he is looking for (line 7): *står der et CVR nummer hernede et sted?* ('is there a CVR number down here somewhere?'). Black's engagement in the interactivity is driven by emotions, frustrations, and the available resources at hand. All in all, Black spends about 20 seconds looking for the CVR number on the noticeboard. This is hardly surprising: in a cognitive ecology, participants look for solutions where they, based on previous experience, think the problem is. At least since Scheerer (1963), this phenomenon has been associated with *fixation bias* and, as well known it rarely leads to successful outcome. Presumably, Black is looking for a CVR number that can be fed into the computer as part of the printed section of the invoice. As Weisberg and Alba (1981, p. 188) note, "one of the difficult aspects of these problems may be that it is not clear that the obvious solutions will not work," and indeed Black's 26 seconds of solution probing cycle did not work. Immediately after the cycle, he finds himself back at the initial point in problem space. A common sense approach to problem solving suggests that if a problem solving strategy fails, another one should be attempted. However, this does not happen: Black embarks on his second cycle of the same strategy, and the repetitive strategy leads to recycling the verbal pattern from line 1: *men jeg kan fortælle dig* (English: 'but I can tell you'), repeated in line 11. Once

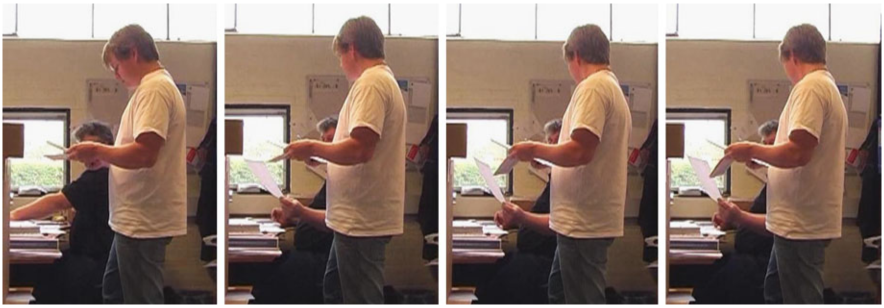


Fig. 11.4 Black picks up the invoice. *Picture 1* (−18150): Black reaches out and grasps the invoice. *Picture 2* (−17025): Black holds the invoice and both participants gaze at it. *Picture 3* (−14125): Black holds the invoice, and the two participants establish eye contact. *Picture 4* (−10425): Black bobs the paper lightly

more, Black recapitulates the problem, and, once more, White does little more than agree (line 14):

```
11. B: men jeg kan fortælle dig, den her den lader de bare ligge.
12.    Den her den betaler de aldrig nogensinde. (3.0)
13.    Den vil aldrig nogensinde blive betalt, den her (1.0)
14. W: Nej nej
```

```
11. B: but I can tell you, this one they'll just discard.
12.    This one they will never ever pay it. (3.0)
13.    It will never ever get paid, this one. (1.0)
14. W: No no
```

However, in complex dynamical systems, there is no such thing as replication of pattern. While the wording may be identical, its physical features and the context of situation can be used by the dialogical system (cf. Steffensen 2012). As shown in Fig. 11.3, the second cycle sets up *enabling conditions* for the transition point at −8575, that is, the problem reframing point. Although unsuccessful in its own terms, the second cycle brings about the interactivity that will later provide the desired solution.

Methodologically, this insight obviously uses a *post hoc* procedure where the salient parts of the third cycle in Fig. 11.3 are traced to the second cycle. So, what are the enabling conditions that lead to a successful third cycle? It will come as no surprise for scholars in distributed cognition that they depend on not only what goes on 'in the head' but also material artefacts. Thus, during the 2175 ms in line 11, Black reconfigures the DCS by means of *artefactual reconfiguration*. Using the invoice, which until this point in time has been lying on the table, Black picks it up and thus establishes a shared focus of attention. This reconfiguration of the DCS exploits the interbodily dynamics of how Black manipulates the artefactual lay-out using his arms and hands. By so doing, Black prompts White's gaze behaviour in a way that contributes to the reconfiguration the DCS (see Ball and Litchfield 2013, Chap. 12). This event is illustrated in Figs. 11.4 and 11.5.

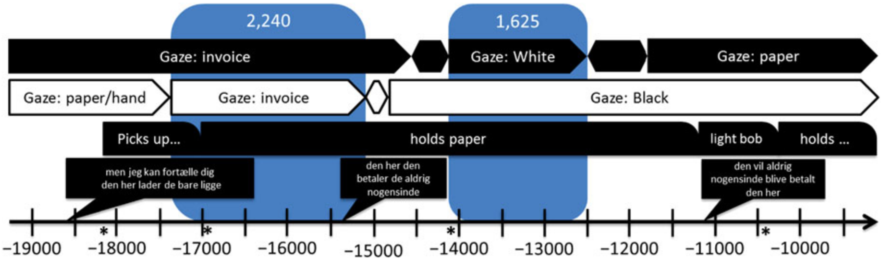


Fig. 11.5 The timeline of the second cycle (from −19000 to −9500). *Black squares* show Black's actions. *White squares* show White's actions. The *two top rows* show gaze direction (*pentagons*) interrupted by head turns (*hexagons*). The *third row* shows Black's hand/arm movements (*squares with a rounded corner*). The *fourth row* shows Blacks wording (in *squared speech bubbles*). The *background* (*blue*) *areas* mark synchronicities in gaze: the shared attention to the invoice (2240 ms) and the eye contact (1625 ms). The *asterisks* at the timeline indicate the four points in time that correspond to the four pictures in Fig. 11.4

As Black picks up the invoice at −18150, his action is synchronised with uttering the demonstrative object deictic *den her* (literally 'this here'; cf. Bang and Døør 2007). This visible and audible movement prompts White to look at the invoice for 2250 ms. In so doing, they establish shared attention to the printed invoice that lies behind the fixation bias. Then, 1000 ms later, they establish eye contact for a remarkably long 1625 ms (from −14125 to −12500, corresponding to the first half of the pause in line 12). Eye contact apparently recalibrates the DCS around the interactivity: they face the same problem. Attention thus subtly shifts from a shared perceptual object to their role as perceptual *subjects*: each becomes aware of the other's role in the problem-solving. Having established this interpersonal frame, Black turns his attention back to the invoice (at −11800). As he does so, he uses a light hand movement to bob it up and down[4] while he utters the initial *den* ('it') in line 13: *Den vil aldrig nogensinde blive betalt, den her* ('It will never ever get paid, this one'). The two participants thus both attend to the invoice and to each other, while concerting the use of the object deictics *den her* ('this one') and *den* ('it') with shifts in gaze and attention.

As part of this second cycle, the events bring about a subtle shift in deictic attention: whereas the first cycle centred on *de* ('they'), an unknown group in their own organisation, the same deictic now identifies the *receiver* of the invoice. In other words, the receiver *lader den ligge* ('just discards it') and thus *aldrig nogensinde betaler den* ('never ever pays it'). The way in which this subtle deictic shift (cf. Steffensen 2012) leads to the secondary event pivot (at −8575), is an example of what Lorenz (1963) famously described as when "slightly differing initial states can evolve into considerably different states," which later became the metaphorical butterfly who by flapping its wings in Brazil sets off a hurricane in Texas. In this

[4]The light shake of the invoice is invisible in the still shots in Fig. 11.4, but a close examination of the 25 frames from the video recording reveals the movement.

case, the “butterfly” is the low-voiced, unnoticeable deictic entrance of the receiver, accompanied by a shift in attention, while the “hurricane” is the identification and, thus, solution of the participants’ real problem.

Re-framing the Problem

From an observer’s perspective the problem is that the printed version of the invoice biases the participants’ thinking. Accordingly, to reach a viable solution or result, they must dissolve their fixation bias. This dissolution occurs in the 8575 ms between two transition points: the problem reframing point and the event pivot. Further, as the terms suggest, they depend on reframing the problem. This event trajectory can be given a first approximation by considering a transcription of Black’s speech:

```
15. B: Hvis det var mig så røg den bare hen i stakken. (0.7)
16.    Den kan jeg ikke betale. (0.4)
17.    Hvorfor kan jeg ikke det? (0.8)
18.    Der er ikke noget CVR nummer på. (0.9)
19.    Du må ikke sende en faktura uden CVR nummer (0.4)

15. B: If it were me then it just went in the pile (0.7)
16.    I can’t pay that. (0.4)
17.    Why can’t I pay it? (0.8)
18.    There is no CVR number on it. (0.9)
19.    You can’t send an invoice without a CVR number. (0.4)
```

In the previous section I argued that the second cycle of the event trajectory brings about the necessary conditions for reframing of the problem, namely: (1) the manipulation of the invoice; (2) the recalibration of the human parts of the DCS, through shared attention to each other and the invoice; and (3) the introduction of the invoice receiver, mediated by the third person deictic *de* (‘they’), and the concurrent shifts in attention. Next, I focus on *how* these three elements dissolve the fixation bias. I argue that they are catalysed by a fourth and novel element which defines the secondary event pivot. The key to understand this element lies in the 1400 ms represented in line 15.

Once again, Black repeats that the problem is that the invoice cannot be paid. However, in this third recycling of the problem, he uses a remarkably different strategy; he does so by attributing a hypothetical *narrative* to the invoice receiver. Linguistically this is marked by a formula for hypothetical thinking: *hvis det var mig* (Eng. ‘if it were me’). Cognitively, the subtle introduction of the receiver in lines 11–12 affects the current trajectory by prompting Black to adopt an *alter-centric* perspective on the invoice. In so doing, he sees the problem from the *perspective of the invoice receiver*. In the tradition of dialogism, Linell (2009, p. 83) refers to this phenomenon as *alter-centric perception*, which he elaborates as follows: “The other’s “outsideness” brings in a ‘surplus’ of vision, knowledge and understanding other than you had before or you had expected to encounter. The other may see things from points-of-view that have so far been strange or unfamiliar to yourself, and this forces you to reflect and try to understand, thereby possibly enriching your, and our collective, knowledge and language".

Linell's description focuses on situations where the two parts of this relation are co-present parties. However, the alter-centric perspective also reflects on non-present, third parties (Linell 2009; cf. Bang and Døør 2007). Indeed, this is why the narrative entails a recalibration of the deictic system, as Black invokes a first person deictic *jeg* ('I') to index the invoice receiver. Thus, whereas in the first cycle (line 12), Black said that *den betaler de aldrig* ('they never pay this'), in this cycle (line 15) the information has been reformulated as *den kan jeg ikke betale* ('I can't pay this'). Such deictic shifts (Steffensen 2012; cf. Bang and Døør 2007) are a cost-efficient way to evoke other parties' perspectives and, in this case, that of the invoice receiver. These deictics "refer directly to the personal, temporal or locational characteristics of the situation" (Crystal 2008, p. 133), and, by so doing, anchor verbal/symbolic dimensions (what has been called "the second order language," cf. Thibault 2011) in real-time speech or first-order languaging. Crucially, then, the result can be traced to neither what is said nor the accompanying non-verbal behaviour: rather it depends on how parties orient to each other with respect to both what is said and how it is that they move. Interactivity, in other words, is sense-saturated in that it links the sense of the deictics (and other words) actually spoken to the pico-scale movements (e.g., shifts in gaze, how syllables are articulated) that are the primordial basis of interactivity.

In short, the DCS uses how deictic markers are integrated with movements and shifts in attention to prompt the participants to reframe the problem. Of course, it would be unwarranted to suppose that the deictics/movements *in themselves* overrule the embodied and situated participant point-of-view. In order to demonstrate the importance of the pico-scale, I examine the 1400 ms of interactivity during which Black utters line 15. This examination, shown in Figs. 11.6 and 11.7, reveals how the participants' interbodily dynamics are enmeshed with the verbal patterns.

As shown in Fig. 11.7, Black's utterance act falls into three intonation groups (here rendered in standard orthography):

hvis det var 'mig	*så røg den 'bare*	*hen i'stakken*
'if it were me'	'then it just went'	'in the pile'

The most prominent syllables *'mig* (['maɪ], 'me'), *'bare* (['bɑː ɑ], 'just'), and *'stakken* (['sdɑgŋ], 'the pile') function as the Zeitgeber of interbodily dynamics. If we are to understand the dissolution of the fixation bias, it is important to attend to how the parties use their timing. Thus, when Black starts his utterance, he is holding the invoice in front of him in his left hand. As he starts on his first intonational group, Black makes a slight upward movement in his left hand; then, he lets go of the paper, and catches it again, app. 3–4 cm closer to the table. The movement is *perfectly synchronised* with the prosodic pattern: the upward movement co-occurs with the first syllable of the utterance, *hvis* ('if'), and the catch of the paper is perfectly synchronised with the prominent syllable *mig*, spoken as his thumb touches

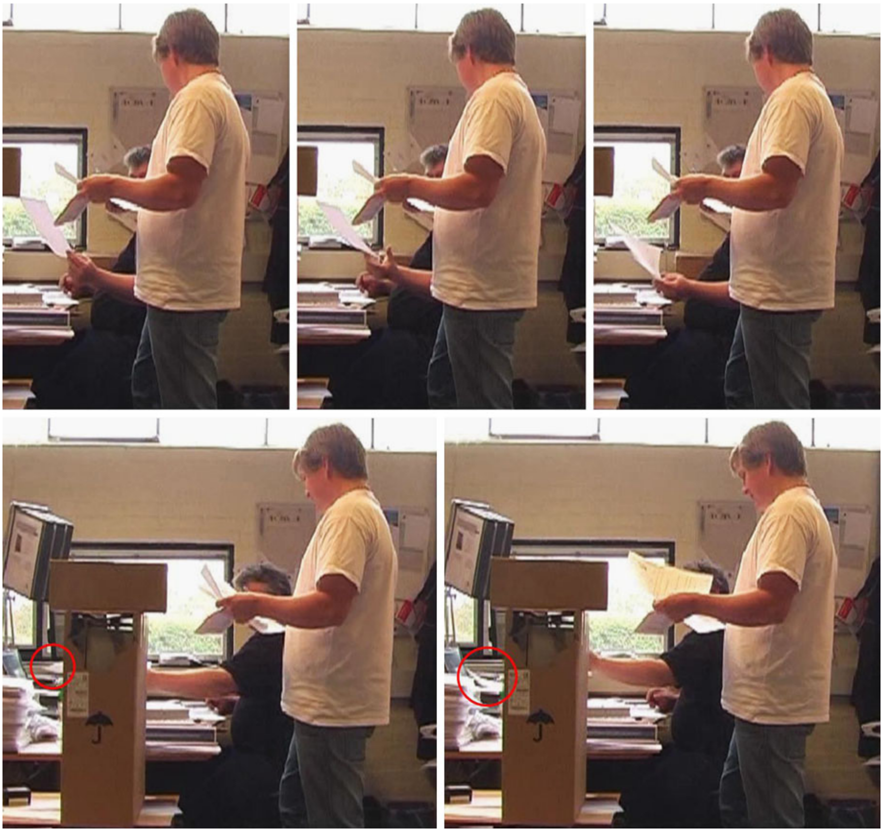


Fig. 11.6 Black embodies the receiver. *Picture 1* (−8500): Black elevates the invoice slightly. *Picture 2* (−8340): Black lets go of the paper and lowers his hand. *Picture 3* (−8140): Black catches the invoice. *Picture 4* (−7140): Black leans forward and stretches his arm so the invoice (indicated by the (*red*) *circle to the left in the picture*) approaches the pile. *Picture 5* (−6940): Black lets go of the invoice (indicated by the (*red*) *circle to the left in the picture*) and lets it drop into the pile

the paper (130 ms into a stressed syllable that lasts 200 ms) just as his voice begins to fall away.[5] In short, as Black adopts an alter-centric deictic system, his body simulates (cf. Kirsh 2013, Chap. 10) a receiver of the invoice. Quite literally, he receives (catches) the invoice at *exactly the same time* as he utters the first person deictic *mig* ('me'). The narrative structure thus prompts him to mimic a receiver in a way that makes his body into a cognitive resource for solving the problem. Black's activity

[5]The margin of error is 40 ms, or the time between two frames. The spectrogram (made in PRAAT software) shows with accuracy when the syllable starts, and this can be imposed on the frame-by-frame video annotation (made in ELAN software). This procedure shows that the hand movement *at latest* starts in the frame that immediately follows the frame in which the syllable starts.

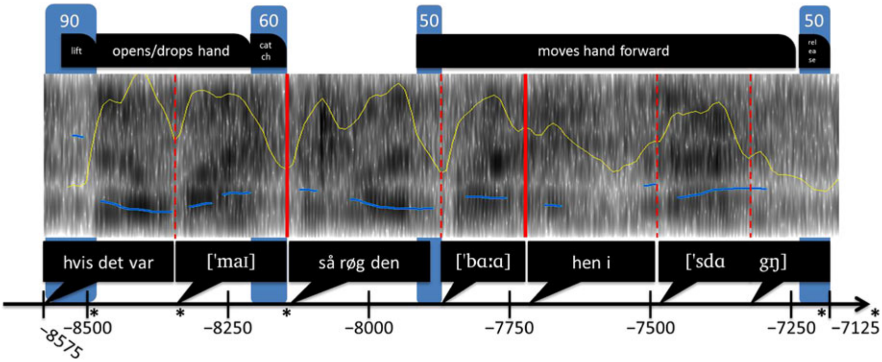


Fig. 11.7 The timeline of line 15 (from −8500 to −7125), corresponding to Fig. 11.6. The *top row* shows Black's hand/arm movements (*squares with a rounded corner*), and the *bottom row* shows Blacks wording with prominent syllables rendered in IPA (in *squared speech bubbles*). Between the two rows, a spectrogram shows some of the acoustic properties of Black's utterance act in line 15, incl. pitch (*blue line* in the *bottom of the spectrogram*) and loudness (*yellow line* in the *top of the spectrogram*). The *background vertical* (*blue*) *areas* mark synchronicities in speech/hand coordination: the beginning of the utterance act and the initiation of the invoice handling (90 ms); the prominent syllable ['maɪ] and the catch of the invoice (60 ms); the prominent syllable ['bɑː ɑ] and the initiation of the forward movement of hand/arm (50 ms); the end of the last syllable [gŋ] and the release of the paper at the end of the forward hand movement (50 ms). The *asterisks* at the timeline indicate the five points in time that correspond to the five pictures in Fig. 11.6. Colour figures can be found in the online version of the chapter

is sense-saturated not only by hypothesising 'if it were me' but also by using deictic recalibration to impose behavioural order.

During the next two intonation groups, Black-as-receiver dissolves the fixation bias based on the print paper version of the invoice. The second intonation group—*så røg den* '*bare* ('then it just went')—pivots on a 150 ms long prominent syllable '*bare* ('just'). As he utters this syllable, his bodily dynamics change: having caught the invoice with his left hand, its downward movement becomes a forward-reaching movement that spreads through his body. He leans forward during the 500 ms of the third intonation group, and as the last syllable ends, lets the paper fall into a pile of paper in the far end of the desk. Remarkably, halfway through Black's *mig* ('me'), White turns his head and follows the invoice with his gaze until Black lets go of the paper at app. −7250. As he does so, White's head movement continues and his gaze rests at the papers in his hand. Black's bodily dynamics set off an interbodily synchrony. This pattern occurs as Black enacts the movement of the imagined invoice receiver who, having realised that he cannot pay the invoice as it carries no CVR number, throws it into a pile of paper; as he does it, White thus observes a scenario of what it looks like when the receiver receives the invoice.

Simulating the receiver has direct implications for the DCS. As argued in Steffensen et al. (2010), a simulation entails a doubling in functional levels of the DCS: it both functions as the *simulation* (where Black plays the role as receiver) and as the *situation* (where Black is himself). Obviously, the discard of the print paper invoice plays out as part of the simulation (it is the imagined receiver who throws it away), but the *effects* of the discard are detectable in the situation (it is Black and White who are no longer biased by the print paper invoice). Thus, the whole-bodied achievement of line 15 recalibrates the DCS which prompts it to solve the problem 7100 ms later. This happens as the two participants move out of the imagined receiver's point-of-view (in lines 15–17) back to their own. Thus, in line 18 Black responds to his own question without personal deictics (*Der er ikke noget CVR nummer på*, 'There is no CVR number on it') and, in line 19, with a deictic *du* ('you'/'one') that is ambiguous between second person and a generalized 'zero' person (cf. Bang and Døør 2007). In short, it is not clear if he is addressing what White can and cannot do or some generic rule of invoice-sending. On either interpretation, Black speaks from an ego-centric perspective that restores his perspective as the *sender* of the invoice. All in all, this third cycle takes the two participants to an alter-centric perspective and back and this cognitive reframing prompts them to dissolve their fixation bias.

Problem-Solving and Solution-Probing

Black has done most of the cognitive labour in the pre-pivot phase. Remarkably, it is White who formulates the solution to the problem in line 20:

```
20. W: nå nej men det er der jo hvis vi printer ud på logo papir. (0.3)
21. B: er der. (0.7)
22. W:↑ja ja (1.2) selvfølgelig (0.7)
23. B: nå ja (0.3) det er rigtigt (0.4)
24. W: ja ja (0.6) den er god nok jo.
25. B: ja det er rigtigt

20. W: well no but it [the number] is there if we print on logo paper. (0.3)
21. B: there is? (0.7)
22. W:↑yes yes (1.2) of course (0.7)
23. B: oh yes (0.2) that's right (0.4)
24. W: yes yes (0.6) it's true, all right (0.4)
25. B: yes that's right
```

Though Black has done the pre-solution work, White now becomes the *main cognizer* (Galosia et al. 2010) by formulating the solution: he says that the number will appear if the invoice is printed on the logo paper. But *how* does he reach the insight or, perhaps, how does the insight reach him? Though no definite conclusion can be reached, it seems clear that this too depends on sense-saturated coordination (as well as neural events). White's bodily actions suggest that the 'insight' arises through four stages as shown in Fig. 11.8: first, the insight comes 'like a bolt from the blue' at −4500; second, he articulates his insight at 0 (the event pivot); third, he realises that the problem has been resolved by the proposed solution (at +4200); and fourth, he outlines consequences of adopting the proposed solution (from +5100 onward).

It may seem surprising to claim that White's insight is manifest 4500 ms before he puts it into words. How do we know that he has reached an insight when he has

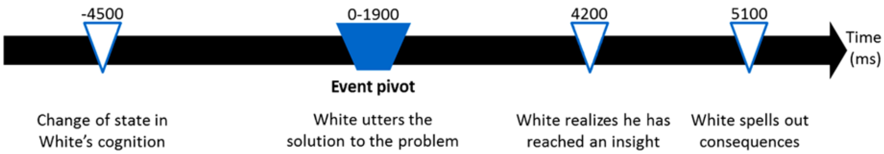


Fig. 11.8 The trajectory of White's insight (from −4500 to +5100). The figure shows the four stages that the DCS undergoes as White reaches the insight

not said so? Close examination of White's posture, head position and gaze, reveals that throughout the whole sequence, from −45000 to −4450, he is in constant motion: he moves slightly back and forth, points to the invoice with the pile of papers in his right hand, fiddles with those in his hand, glancing occasionally at them as his gaze follows the invoice in Black's hand. But at −4450 White suddenly looks up from his paper, gazes towards Black's computer screen, and, during Black's utterance act in lines 18–19, stands completely motionless for 4000 ms. Though we cannot observe any cognitive work, we can assume that *neurally* something is happening. Briefly, the DCS pivots on a brain that comes up with the logo paper hypothesis. This change occurs 560 ms after Black ends his alter-centric dissolution of the fixation bias (in lines 15–17): this short time frame indicates a direct link between Black's narrative and White's insight.

As White utters the insight in line 20, he turns his head and gazes at Black's computer while pointing at the screen. Strikingly, White gives no sign of noticing that he has solved the problem. The first indication appears when White, at +3230, responds to Black's *er der?* ('there is?'). He now uses a rising intonation on *ja ja* ('yes yes', line 22) that, by +3800, merges into a clear indication that he has grasped the implications of what he has said. White closes the pile of papers in his hand, looks up, and leans forward as he stretches his left arm to pick up the invoice from the pile at Black's table. In the midst of the bodily movement, as he utters the prominent syllable [ˈfø] in *selvfølgelig* ([sɛˈføli] 'of course', line 22), he hesitates for 400 motionless ms. Again, it seems that the DCS has come to pivot on White's neural resources as he grasps the implications of a solution.

If it is indeed the case that White utters the solution at 0, and realises that he had done so at +4200, it implies that the two participants did not exploit language as a tool for aligning states of mind through externalisation, as assumed by mainstream linguists and cognitive scientists. Rather, the wording in line 20 functions as a cognitive probe beyond deductive logic and problem-solving strategies. Thus, rather than engaging in problem-solving, the two participants depend on *solution-probing*: they project probes until they, *post festum*, observe that one of the probes fits the problem space. The participants are thus not merely agents, but also observers of their own interactivity. It is this probing strategy that gives the cognitive trajectory its chaotic, self-organising quality: it does not follow a pre-destined scheme or blueprint; it is messy and meshed with real-time interbodily behaviour.

The post-pivot period is characterised by White's refocusing attention on the invoice which he grasps and places in front of Black. As they have now dissolved the fixation bias, the printed invoice no longer represents their impasse; it is, rather,

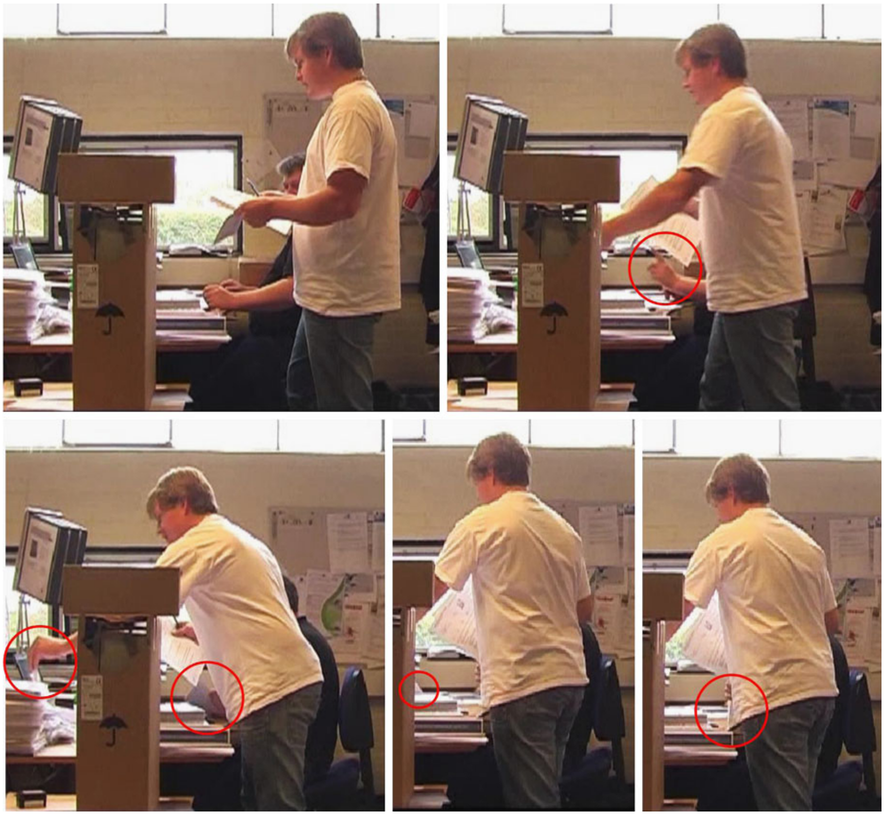


Fig. 11.9 Black and White synchronises a pick-up of papers. *Picture 1* (+5050): Immediately prior to the sequence. *Picture 2* (+5650): White has started leaning forward, reaching for the invoice in the pile; Black lifts his index finger (marked by *circle*). *Picture 3* (+7010): White has grasped the invoice and starts lifting it (marked by *circle*); Black has grasped the papers at the table and starts lifting them (marked by *circle*). *Picture 4* (+8130): White lowers the invoice which is just about to touch the table (marked by *circle*). *Picture 5* (+8290): The papers in Black's hand (marked by *circle*) have just made contact with the table, 100 ms after the invoice in White's hand did the same

a resource for completing a workable invoice. In the post-pivot period, they undertake excessive confirmations (lines 22–25) that align the two participants in positive emotional resonance that contrasts with the pre-pivot frustration. They resonate not only through verbal affirmations but also in pico-scale interbodily dynamics. This resonance occurs when White picks up the paper and places it on Black's table. As he does so, Black synchronises with White's movement by picking up a pile of papers, and stomping them in the table. Not only is the movement similar: both participants lift their respective papers from a horizontal position to touch the table in a vertical one. What is most striking is that the movements are almost completely synchronous, as shown in Figs. 11.9 and 11.10.

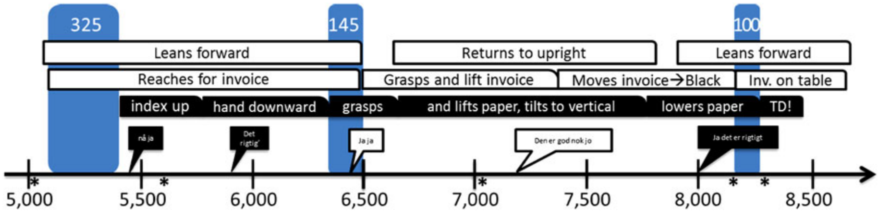


Fig. 11.10 The timeline of the interbodily synchronisation in Fig. 11.9 (from +5000 to +8600). The four rows show: White's posture, White's arm/hand, Black's arm/hand, and Black's and White's speech. The synchronisation is marked with *rounded squares in the background* (in *blue* in the electronic version) first when the initiate the hand movements (within 325 ms), second when they pick up the invoice/the papers (within 145 ms), and third when they make the invoice/the papers touch the table (within 100 ms). The *asterisks* at the timeline indicate the five points in time that correspond to the four pictures in Fig. 11.6

Black's hand movement starts 325 ms after White's, and strikingly, he grasps his pile of papers 145 ms before White. As Black's hand is closer to his paper than is White's to the invoice, Black makes a circuitous movement: he lifts his hand and points his index finger upwards (in a *eureka*-like way), that is synchronised with a *nå ja* ('oh yes') in line 23 before he proceeds by lowering his hand towards the papers. Within 145 ms both participants lift their respective sheets, move them in an arc, and make the bottom edge of the paper touch the table within 100 ms of each other. On the pico-scale of interbodily dynamics, they engage in a spontaneously choreographed ballet where their papers are moved in a resonant, synchronous pattern that resonates with their (by Danish standards) excessive verbal agreement.

Conclusion: Interactivity, Language and Cognition

In the previous section we have undeniably seen a problem being solved. Loosely speaking, we could say that Black and White solved the problem, more technically that it was solved by a DCS. In terms of the components hierarchy suggested above, the DCS under scrutiny included human beings, artefacts, and narratives. Obviously, a large part of the cognitive labour consisted in how the interbodily dynamics of the two participants iteratively calibrated and aligned the DCS. These dynamics comprised: (i) Sharing of emotional states (e.g., frustration before the event pivot and exultation after); (ii) timing of vocal gestures (e.g., syllabic prominence as a Zeitgeber for bodily actions); (iii) coordination of gaze: partly as shared attention at the invoice, partly as eye contact; (iv) synchronisation of movement (e.g., in how papers are handled).

Especially the handling of the print paper invoice played a vital part in the cognitive event trajectory. Thus, the invoice became a main artefact for Black as he reframed the problem through a simulation of how the receiver would handle the invoice. Significantly, the invoice undertook a spatial reorganisation, as the two participants actively turned their environment into a cognitive resource. As indicated in

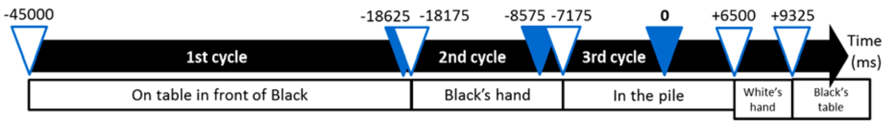


Fig. 11.11 The trajectory of the print paper invoice throughout the cognitive event. The figure shows the spatial reorganisation from the table to the pile and back, mediated by Black's and White's handling it, respectively

Fig. 11.11, the spatial reorganisation correlated with the event trajectory: the invoice both starts (until −18625) and ends (from +6500) on the table in front of Black, first as a fixation bias, then as a resource in their future work. Between these two points, the invoice is placed in the pile at the far end of the table. Strikingly, Black puts it there, while White takes it back.

The active, real-time, spatial manipulation of artefacts thus contributes to the cognitive event trajectory. One could argue that the spatiality of the DCS environment is thus a main component that should be added to the hierarchy established earlier in this chapter. This would be in line with Hollan et al.'s (2000:179) observation: "To understand human cognition, it is not enough to know how the mind processes information. It is also necessary to know how the information to be processed is arranged in the material and social world." The importance of the spatial dimensions also came to the fore when Black scanned his noticeboard as part of the environment in search for the CVR number.

The narrative contributed to the DCS through its ability to prompt the participants to see the problem from other points of view, *in casu* the view of the receiver of the invoice. This was achieved partly through the narrative formulaic *hvis det var mig* ('if it were me'), partly through the use of personal deictics grounded in an alter-centric perspective. All in all, the wording and the narrative prompted the participants to reframe the cognitive problem at hand.

The analysis also gave rise to another insight into the nature of problem-solving. Far from being a testimony of the cognitive powers of human rationality, the analysis showed a modus operandi that does not depend on developed schemes, plans, and blueprints for solving problems. The DCS under scrutiny moved through the problem-space along a cognitive trajectory that was self-organised, unpredictable and on the edge of chaos. We did not witness analytical problem-solving, but rather creative *solution-probing*. The probing nature of how Black and White solved the problem points in two directions. First, it seems that the pervasiveness of social institutions that operate in "highly rationalised environments" (McGarry 2005, p. 187)—including science, at least as it presents itself front-stage—has biased our understanding of how human beings solve problems. Such institutions depend on standardised and automatised modes of working where norms and procedures impose an orderliness that turns creative solution-probing into analytical problem-solving. However, even in such rationalised environments, the DCS has an underlying capacity for self-organising cognitive processes in a way that phenomenologically appears to us as *creativity*, *intuition*, and *Aha!* Second, it seems that standard versions of theories on human problem-solving depend on analytical and cognitive psychological models of processing input (the problem) in order to generate output (the solution). This calls for a theoretical reassessment of the embodied and interbodily dynamics of human problem-solving, a reassessment that upholds the procedural model of problem-solving as a subset of a more general human capacity. Such a subset comes in handy in complex work environments where speed or security become decisive, for example in emergency medicine (cf. Pedersen 2012) or in aviation (cf. Desai et al. 2012), but even in such environments intuition and creativity can be crucial elements in achieving a successful result.

This chapter has made two contributions to a more realistic view on human problem-solving. First, it devised a method of *Cognitive Event Analysis*. Most saliently, this method pivots on scrupulously close examination of video-recorded events (building on such work as Cowley 1998; Steffensen et al. 2010; Thibault 2011; Pedersen 2012). The method thus allows for detailed scrutiny of *what happens* as cognitive results are brought forth. It consists of two steps, an event identification procedure, and a cognitive event trajectory analysis. It thus combines a deductive procedure with an inductive approach to data. While the former allows the method to integrate in various trans-disciplinary fields (e.g., the study of cognitive events in the workplace, in aviation, or in health settings), the latter appreciates Clark's (2008, p. 13) *Principle of Ecological Assembly*. Thus, if the enabling conditions of problem-solving include "whatever mix of problem-solving resources [that] will yield an acceptable result," there is no *a priori* way of delineating what ought be to investigated as contributing to the emergence of a solution. As argued, the whole array of situated elements, including interbodily dynamics and artefacts, may contribute to the cognitive event, but the same goes for many non-situated elements, including sociocultural resources, verbal patterns, narratives, memorised procedures, and autobiographical memory. Such non-situated components of a DCS pose a methodological problem for observational methods, because they depend on the non-situated parts being situated, and that does not always happen in an overt way. However, if the observational data are complemented by careful ethnographic methods—for example, interviewing and participant observation—it is to an extent possible to counter this shortcoming. Likewise, a careful inclusion of experimental data may prove fruitful: Thus, if a given structure is known to yield cognitive results in controlled experimental settings, the same structure is, *ceteris paribus*, likely to yield similar results when appearing in the wild.

The second contribution of the current chapter consisted in an interactivity-based approach to problem-solving. The present analysis points to the necessity of grounding the study of cognition in interactivity, or sense-saturated coordination that contributes to human action. In allowing interactivity to play the decisive role in human problem-solving, we remove some of the cognitive burden from the brain. This view on cognition contrasts with what we can term the "dermal metaphysics" of traditional cognitive science, sociology, and biology, i.e. the view that the skin is the absolute barrier between two distinct realms: the internal vs. the external, biology vs. sociality, cognition vs. communication, or meaning vs. behaviour. With interactivity, we need not make any such a priori distinctions: They all come together in what people do as they achieve results. Thus, interactivity is an ontological substrate that can be described both as cognition, as language, as ecological niche-construction, and as behaviour. Accordingly, we must discard "(T)he simplemindedness of mainstream cognitive psychology about the nature and function of human communication" (Reed 1996, p. 157): language is *not* an instrument for externalising thoughts, language is *not* an instrument for exchanging pre-existing meaning, and language is *not* a pre-existing system that we can "use." Rather, language is meshed with real-time interbodily behaviour. It functions as a Zeitgeber for action (cf. Steffensen et al. 2010) and as a constraint on pico-scale behaviour. The latter is exemplified by Black and White when they exploited the symbolic dimensions of language to simulate the receiver's point of view. In this way, language enables participants to perform low-cost categorisation and simulation in their ecology (cf. Clark 2008, Chap. 3). The mechanism that allows us to do this is our capability to take a *language stance* (Cowley 2011). The language stance allows us to control our pico-scale bodily behaviour (e.g., as we produce and perceive specific vocal gestures) in a way that conforms to our phenomenological experience of engaging with verbal patterns.

Finally, it should be pointed out that an interesting implication of an interactivity-based approach is that it links to the philosophical program of Ludwig Wittgenstein (1953). Just as Wittgenstein sought to move beyond, or below, linguistic meaning with recourse to what people do with language, the interactivity approach seeks to move below the phenomenology of language by exploring human behaviour on a pico-scale. On the micro-scale of wording and social interaction, people can agree (or not): they can align verbally, phenomenologically and consciously. On the pico-scale, people can resonate and synchronise (or not)—they can align sub-consciously, sub-phenomenologically—or just interbodily. With Wittgenstein, we can say that on the pico-scale there "is not agreement in opinions but in form of life" (Wittgenstein 1953:88 [§241]).

Acknowledgements I am grateful to the editors of this volume for the meticulous comments and suggestions, and not least their Zen-like patience. Likewise, Sarah Bro Pedersen (University of Southern Denmark) contributed with helpful suggestions, and Rasmus Nielsen (University of Southern Denmark) assisted in the interpretation of the spectrogram in Fig. 11.7. Emo Gotsbachner (University of Vienna) contributed with critical comments that sharpened my interpretation of the event pivot. Finally, I am indebted to Anne-Kathrine Brorsen who allowed me to use her data for the in-depth analysis, and not least indebted to Black and White for sharing their problem-solving with us.

References

Ball, L. J., & Litchfield, D. (2013). Interactivity and embodied cues in problem solving, learning and insight: further contributions to a "theory of hints". In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Bang, J. C., & Døør, J. (2007). *Language, ecology and society: a dialectical approach*. Edited by S. V. Steffensen and J. Nash. London: Continuum.

Berkowitz, L. (2007). Frustration-aggression hypothesis. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of social psychology* (pp. 366–368). Thousand Oaks: Sage.
Brorsen, A.-K. (2010). *Til dialog…* Unpublished elite module project. Odense: University of Southern Denmark.
Bråten, S. (2009). *The intersubjective mirror in infant learning and evolution of speech*. Amsterdam: Benjamins.
Chemero, A. (2000). What events are: commentary on Stoffregen's "Affordances and events". *Ecological Psychology*, *12*, 37–42.
Clark, A. (2008). *Supersizing the mind: embodiment, action and cognitive extension*. Oxford: Oxford University Press.
Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7–19.
Cowley, S. J. (1998). Of timing, turn-taking, and conversation. *Journal of Psycholinguistic Research*, *27*, 541–571.
Cowley, S. J. (2011). Distributed language. In S. J. Cowley (Ed.), *Distributed language* (pp 1–14). Amsterdam: Benjamins.
Cowley, S. J. (2012). Mimesis and language: a distributed view. *Interaction Studies*, *13*, 17–40.
Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th ed.). Oxford: Blackwell Sci.
Desai, V., Madsen, P. M., & Roberts, K. H. (2012). High-reliability organizations in health care. In P. Carayon (Ed.), *Handbook of human factors and ergonomics in health care and patient safety*. Boca Raton: CRC Press, Taylor & Francis.
de Saussure, F. (1916/1972). *Cours de linguistique générale*. Paris: Payot.
Dollard, D., Doob, L., Miller, N., Mowrer, O., & Sears, R. (1939). *Frustration and aggression*. New Haven: Yale University Press.
Donald, M. (1991). *The origins of the modern mind*. Cambridge: Harvard University Press.
Donald, M. (2001). *A mind so rare*. New York: Norton.
Donald, M. (2012). Commentary on Cowley's chapter. *Interaction Studies*, *13*, 41–49.
Galosia, M., Cowley, S. J., Aliner, G., & Steffensen, S. V. (2010). When life hangs on biosemiosis: tracking expertise in a medical simulation centre. In S. J. Cowley, J. C. Major, S. V. Steffensen, & A. Dinis (Eds.), *Signifying bodies: biosemiosis, interaction and health* (pp. 173–200). Braga: Portuguese Catholic University Press.
Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, *32*, 1489–1522.
Harel, I. (1991). The silent observer and holistic note taker: using video to document a research project. In I. Harel & S. Papert (Eds.), *Constructionism*. Norwood: Ablex.
Hemmingsen, B. (2013). The quick and the dead: on temporality and human agency. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.
Hodges, B. H., & Baron, R. M. (1992). Values as constraints on affordances: perceiving and acting properly. *Journal for the Theory of Social Behaviour*, *22*, 263–294.
Hollan, J., Hutchins, E., & Kirsch, D. (2000). Distributed cognition: toward a new foundation for human–computer interaction. *ACM Transactions on Computer-Human Interaction*, *7*, 174–196.
Hutchins, E. (1995a). *Cognition in the wild*. Cambridge: MIT Press.
Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science*, *19*, 265–288.
Järvilehto, T. (1998). The theory of the organism-environment system: I. Description of the theory. *Integrative Physiological and Behavioral Science*, *33*, 321–334.
Järvilehto, T. (2009). The theory of the organism-environment system as a basis of experimental work in psychology. *Ecological Psychology*, *21*, 112–120.
Kirsh, D. (1997). Interactivity and multimedia interfaces. *Instructional Science*, *25*, 79–96.
Kirsh, D. (2006). Distributed cognition: a methodological note. *Pragmatics & Cognition*, *14*, 249–262.
Kirsh, D. (2013). Thinking with external representations. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Linell, P. (2009). *Rethinking language, mind, and world dialogically: interactional and contextual theories of human sense-making*. Charlotte: Information Age Publishing.
Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*, 130–141.
Maier, N. R. F. (1930). *Problem solving and creativity in individuals and groups*. Belmont: Brooks.
McGarry, B. (2005). *Things to think with: understanding interactions with artefacts in engineering design*. Ph.D. Dissertation, University of Queensland, St Lucia.
Pedersen, S. B. (2010). Towards dialogical health care practices: human errors as a result of cultural affordances. In S. J. Cowley, J. C. Major, S. V. Steffensen, & A. Dinis (Eds.), *Signifying bodies: biosemiosis, interaction and health* (pp. 245–276). Braga: Catholic University of Portugal.
Pedersen, S. B. (2012). Interactivity in health care: bodies, values and dynamics. *Language Sciences*, *34*, 532–542.
Perry, M. (2013). Socially distributed cognition in loosely coupled systems. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.
Reed, E. S. (1996). *Encountering the world: toward an ecological psychology*. Oxford: Oxford University Press.
Robbins, P. & Aydede, M. (Eds.) (2009). *The Cambridge handbook of situated cognition*. Cambridge: Cambridge University Press.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735.
Scheerer, M. (1963). Problem solving. *Scientific American*, *208*, 118–128.
Steffensen, S. V. (2009). Language, languaging and the extended mind hypothesis. *Pragmatics & Cognition*, *17*, 677–697.
Steffensen, S. V. (2011). Beyond mind: an extended ecology of languaging. In S. J. Cowley (Ed.), *Distributed language* (pp. 185–210). Amsterdam: Benjamins.
Steffensen, S. V. (2012). Care and conversing in dialogical systems. *Language Sciences*, *34*, 513–531.
Steffensen, S. V., & Cowley, S. J. (2010). Signifying bodies and health: a non-local aftermath. In S. J. Cowley, J. C. Major, S. V. Steffensen, & A. Dinis (Eds.), *Signifying bodies: biosemiosis, interaction and health* (pp. 331–356). Braga: Catholic University of Portugal.
Steffensen, S. V., Thibault, P. J., & Cowley, S. J. (2010). Living in the social meshwork: the case of health interaction. In S. J. Cowley, J. C. Major, S. V. Steffensen, & A. Dinis (Eds.), *Signifying bodies: biosemiosis, interaction and health* (pp. 207–244). Braga: Catholic University of Portugal.
Thibault, P. J. (2011). First order languaging dynamics and second-order language: the distributed language view. *Ecological Psychology*, *23*, 210–245.
Trevarthen, C. (1998). The concept and foundations of intersubjectivity. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 15–46). Cambridge: Cambridge University Press.
Vallée-Tourangeau, F., Euden, G., & Hearn, V. (2011). Einstellung defused: interactivity and mental set. *Quarterly Journal of Experimental Psychology*, *64*, 1889–1895.
Weisberg, R. W., & Alba, J. W. (1981). An examination of the alleged role of "fixation" in the solution of several "insight" problems. *Journal of Experimental Psychology. General*, *110*, 169–192.
Wertsch, J. (2002). *Voices of collective remembering*. Cambridge: Cambridge University Press.
Wilson, R. A., & Clark, A. (2009). How to situate cognition. letting nature take its course. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 55–77). Cambridge: Cambridge University Press.
Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell Sci.

Chapter 12
Interactivity and Embodied Cues in Problem Solving, Learning and Insight: Further Contributions to a "Theory of Hints"

Linden J. Ball and Damien Litchfield

Abstract This chapter addresses the situated, embodied and interactive characteristics of problem solving by focusing on the cues that arise within a solver's external environment. In examining the influence of external cues on problem solving we have been heavily influenced by Kirsh's (The Cambridge handbook of situated cognition, Cambridge University Press, Cambridge, 2009) "theory of hints". We extend this theory to include hints that derive from the communicative properties of other people's eye movements, focusing on the role of eye gaze in directing attention and conveying information that can be beneficial for problem solving. A particularly interesting aspect of eye gaze is its capacity to facilitate the perceptual priming of motor simulations in an observer. This gives rise to the potential for an expert problem solver's eye movements to cue imitative perceptual and attentional processing in less expert observers that can promote effective problem solving. We review studies that support the hypothesised role of gaze cues in scaffolding problem solving, focusing on examples from insight tasks and diagnostic radiography. Findings reveal that eye gaze can support a variety of decisions and judgements in problem solving contexts. In sum, knowing where another person looks provides hints that can act both implicitly and explicitly to cue attention and to shape thoughts and decisions.

There are many occasions when we find that we have a problem to solve, that is, we have a goal that we want to achieve but do not immediately know what to do to reach that goal. Such problems can be fairly mundane, such as trying to keep dry when you get caught in an unexpected downpour or trying to find your way around a new city that you are visiting for the first time. The problem continuum can also stretch to more profound goals, such as working out what you need to do to avoid

L.J. Ball (✉)
School of Psychology, University of Central Lancashire, Darwin Building, Preston, PR1 2HE, UK
e-mail: LBall@uclan.ac.uk

D. Litchfield
Department of Psychology, Edge Hill University, St Helens Road, Ormskirk, Lancashire, L39 4QP, UK
e-mail: damien.litchfield@edgehill.ac.uk

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_12, © Springer-Verlag London 2013

bankruptcy or striving to resolve a major scientific research question. An issue that is intimately connected to problem solving is that of “transfer”, which is primarily concerned with the benefits that prior experience and knowledge can bring to current problem solving. The transfer theme connects very closely with topics such as learning and the development of expertise, since there is much research showing that as individuals engage deliberatively in tackling problems within a particular domain they transition from novice to expert status through processes of knowledge acquisition and knowledge restructuring (e.g., Van De Weil et al. 2000) as well as through the development of highly effective storage and retrieval mechanisms to support the use of domain-specific knowledge (e.g., Ericsson and Kintsch 1995). Indeed, when advanced levels of expertise are attained many tasks within a domain may end up becoming fairly routine, no longer having the status of traditional problems, since the experienced practitioner not only possesses the requisite knowledge to solve the task but also possesses the methods that are needed to ensure the effective retrieval and application of such knowledge.

In this chapter we begin by considering some of the changing theoretical conceptions of problem solving and learning that have arisen from a growing appreciation that these activities need to be understood as being fully situated, embodied and interactive in nature (e.g., Kirsh 2009), as opposed to the classical view (e.g., Newell and Simon 1972), whereby problem solving is decontextualised, disembodied and divorced from external resources that are present within the solver’s physical and cultural environment. The present chapter will touch upon a variety of important issues that relate to the situated, embodied and interactive characteristics of problem solving as conceptualised in contemporary research, with a key focus throughout our discussion being placed on the crucial role played by cues that arise within the solver’s external environment.

In this latter respect we agree fully with Kirsh’s (2009) assessment that classical problem solving theory has failed adequately to grapple with “the universality of cultural products that facilitate activity-specific reasoning” (Kirsh 2009, p. 284). What Kirsh is referring to here is the idea that the environments in which people regularly act are replete with “mental aids”, such that problem solving becomes more a matter of making effective use of these available aids than relying purely on existing knowledge and internal cognition. As Kirsh goes on to say, “Our problems arise in socially organised activities in which our decisions and activity are supported” (p. 284). Such support structures take the form of “scaffolds” and “resources” that are designed to make it easier for people to complete their tasks, whether these are wayfinding problems, problems in using technological devices and appliances or problems in choosing goods in a local supermarket to meet nutritional preferences and budgetary constraints. Scaffolds and resources can, more generally, be viewed as “hints”, with a key source of such hints being our colleagues, neighbours, supervisors, teachers, trainers and the like, who are usually at hand to offer assistance by providing helpful suggestions, clues, advice and tools. Kirsh (2009) views such hints as providing a key basis for an alternative and positive theory of how people overcome problems in concrete settings through a dynamic process of agent-environment interaction.

We have taken inspiration from Kirsh's (2009) recent sketch of a "theory of hints", and in the present chapter we attempt to provide a small addition to this outline theory, albeit an addition that we believe is a vitally important one. Our particular focus is on the role of hints that derive from the communicative properties of others' eye movements in directing attention and conveying information (e.g., Kleinke 1986). The question underpinning our research is a simple one: if a person's eye movements can direct our attention, then can we also learn to make use of these cues wherever possible in a way than can shape and facilitate our problem solving? This is an important empirical question that we and other contemporary researchers are actively addressing, and one of the major goals of this chapter is to review some of the key literature in this area. By way of pre-empting our answer to the question, we note up-front that the existing evidence points resoundingly to an affirmative conclusion. We further suggest that such a positive answer also brings with it the need to augment a theory of hints in a way that accommodates the problem solving scaffolds than come from our inherent human sensitivity to other people's eye gaze.

Changing Conceptions of Problem Solving and Learning

The core focus of much traditional research on problem solving, learning and expertise up to the 1990s was on relatively small-scale tasks being undertaken by individuals working within highly controlled laboratory conditions. Well-defined "puzzle" problems (e.g., the Tower of Hanoi task and river-crossing problems) and games (e.g., chess and tic-tac-toe) featured heavily in this research endeavour and formed the "fruit-flies" of much influential early theorising, inspiring the classical theory of problem solving espoused by Newell and Simon (e.g., 1972; see also Simon 1981). This classical theory views problem solving as involving a heuristically-guided search through a representation of the task (i.e., a "problem space") from an initial state to a goal state via intermediate states. This classical view has many strengths, not least the elegance of portraying a highly generalised view of human problem solving as an adaptive process that is finely tuned to a set of environmental constraints—the so called "task environment" of a problem—that can be conceptualised by theorists as forming the core, abstract, structural aspects of the problem. An individual's problem space may represent more or less of this task environment, since there may be omissions or commissions, the latter deriving from prior biasing assumptions that the individual brings to bear. In addition, the problem space may not only comprise a mental representation but may also be distributed over external resources (e.g., Larkin and Simon 1987; Tabachneck-Schijf et al. 1997).

Even during the heyday of this classical approach to explaining problem solving there were many dissenting voices, which came from those who were concerned that studying decontextualised problem solving by lone individuals was missing much of the richness of real-world problem solving. Those clamouring for a re-focusing of the research agenda tended to base their arguments on two major issues. First, the nature of most real-world, problem solving activity means that it is highly "situated", such that cognitive processes are likely to be shaped heavily by organisational and cultural goals, social structures and interpersonal interactions, as well as by the external artefacts, tools and representational systems at people's disposal (e.g., Suchman 1987). As an example, take the problem solving that arises in the domain of professional design practice. Commercial design is typically observed to be heavily bounded by particular company contexts, which means that design processes are influenced by the constraints and affordances that derive from team members and managers, organisational priorities and goals and established cultural conventions (e.g., see Ball and Ormerod 2000a, 2000b; Reid et al. 2000).

Second, real-world problem solving tends to be highly "distributed", in that cognition it is not located within any one individual, but is instead mediated through complex interactivity between multiple internal and external knowledge repositories, including other team members and various external artefacts (e.g., Busby 2001; Lave 1988). One upshot of this "distributed cognition" approach (e.g., Hutchins 1995) is that it brings into sharp focus the inherent poverty of the classical view, in which context and culture merely moderate the internal cognitive processes of individuals. Instead, it is proposed that "cultural activity systems" (Hutchins 1995) can have cognitive properties of their own that reflect emergent aspects of the whole, rather than simply being the sum of the properties of the individuals working within the system.

Somewhat paradoxically, much of the success of the classical theory of problem solving arguably derived by virtue of the way in which context and culture were effectively *controlled out* of laboratory-based experiments, in essence giving a sterilised view of the reality of real-world problem solving As a consequence, what subsequent problem solving research has revealed very pointedly is that there can be major discrepancies between phenomena that arise "in vivo" and phenomena that have been established "in vitro". For example, Dunbar and Blanchette (2001) demonstrate that the wealth of spontaneous analogical reasoning and communication that arises in real-world problem solving all but disappears in laboratory studies, unless people are explicitly instructed to analogise. In a similar vein, Hutchins (1995) observed how the well-established laboratory-based phenomenon of "confirmation bias" (i.e., a tendency for *individuals* to engage in verifying rather than falsifying tests of favoured hypotheses) was absent in team-based hypothesis-testing in a navigational decision-making context. Like Hutchins, we have also provided evidence for another well-known cognitive tendency, "satisficing"—where problem solvers fixate upon a satisfactory solution rather than exploring options to go beyond mere satisfactory outcomes—as being a dominant force in individual design activity, while being largely eradicated in the interactivity arising in team-based design practice (Ball and Ormerod 2000b).

Nowadays, few researchers remain wedded to the classical theory of problem solving. Instead, much research attention is now focused on developing theories that fully embrace the situated, embodied and interactive nature of problem solving using methodologies such as cognitive ethnography that are appropriate for capturing and analysing the richness of real-world cognition (see Ormerod and Ball 2007). As we noted above, our particular aim in the present chapter is to focus on the interactivity that arises in real-world, distributed problem solving contexts in order to contribute further conceptual insights to Kirsh's (2009) recent sketch of a "theory of hints". Hints take the form of a wide variety of scaffolds and resources that often originate from the people around us (e.g., colleagues, managers, advisors and the like), who provide clues, suggestions and tools. As Kokinov et al. (1997) note, hints provide a crucially important element of the "culture" of problem solving. Given their apparent importance, we agree with Kirsh's (2009) proposal that any theory of situated problem solving also needs to explain why hints are so successful. Likewise, such a theory needs to be able to accommodate the many different ways in which our problem solving environments offer up hints that enable us to tackle problems more effectively.

In outlining what a relatively simple theory of hints might look like, Kirsh (2009) suggests that it should begin by first defining what a hint is, which he takes to be a verbal or nonverbal cue that acts like a heuristic that can bias the search process. This definition is appealing, not least because it aligns in a constructive way with key concepts from classical problem solving theory, whereby problem solving is viewed as heuristically-guided search though a problem space. Of course, contemporary situated accounts of problem solving downplay the notion of problem-space search in favour of concerns with the external environment and the socio-cultural context as well as with the way in which such factors ground cognition. But as Kirsh emphasises, no comprehensive account of problem solving can overlook the vital importance of understanding the way in which candidate solution ideas are "generated" and subsequently "evaluated" by the problem solver. In this way, an analysis of the manner in which hints provide candidate solutions whose adequacy can be tested gives a foundation for something of a rapprochement between the classical and situated accounts of problem solving. Indeed, the upshot of such a reconciliation of views is that the resulting theory has the potential to afford a highly positive account of how people overcome problems in concrete settings through a dynamic process of agent-environment interaction.

Kirsh's theory views hints as having a particularly valuable *generative* function in games such as chess, where there are a vast number of choice points that need to be considered to make an effective move. Thus, typical verbal hints for opening a game include advice such as "Open with a centre pawn" or "Knights before bishops", which serve to bias candidate generation to prudent options. In standard, well-defined "puzzle" tasks, where only a relatively small number of discrete moves are possible, Kirsh suggests that hints are more likely to have an *evaluative* function, helping the problem solver to determine whether a move or action is a good one that has the potential to lead to the desired goal state. For other types of problems, such as those where it is not easy to determine what the available options are or even whether one is at a choice point, Kirsh suggests that hints are most likely to be beneficial if they can help one to *frame* the problem in a constructive manner. In this way, hints may help the problem solver to break away from problem frames arising from the inappropriate application of prior assumptions and beliefs. Such false assumptions may result in the problem solver succumbing to mental "set" or "fixation", which, in turn, can induce a phase of "impasse", where the individual is stuck and cannot make further progress. Indeed, problems that are very hard to solve tend to be of this type, where people find it difficult to escape the shackles of familiar ways of proceeding that are, in fact, inappropriate for solving the current task (Dominowski and Dallob 1995). Such problems are referred to as "insight" tasks (e.g., Gilhooly and Murphy 2005; Kaplan and Simon 1990; Ohlsson 1992), since a successful solution typically only arises as a consequence of a radical reframing or restructuring of the problem that engenders a sudden insight as to the necessary path to a solution (Smith and Kounios 1996).

Hints in Insight Problem Solving

In the case of insight problem solving, pioneering studies by Gestalt psychologists have shown how environmental hints that direct attention to a particular item of information associated with the problem setting can assist the problem solver in finding an appropriate problem frame that can lead to an insightful solution. A classic case of the benefits of such environmental hints can be seen in Maier's (1931) study of the two-string problem. In his study, Maier brought the participant into a room containing various objects (e.g., poles, pliers) as well as two strings that were hanging from the ceiling. The participant was then asked to tie together the two strings, but in attempting to do this discovered that the strings were too far apart, such that it was impossible to reach one string while holding onto the other one. The "insight" solution—provided by only a few participants—was to tie the pliers to one of the strings and set it swinging like a pendulum so as to be able to catch it on the upswing while holding the other string. Crucially, Maier found that this insight could be engendered spontaneously by having the experimenter accidentally brush against one of the strings so as to set it swinging, thereby providing a subtle hint that the participant could exploit. Maier argued that participants were not consciously aware of being influenced by the experimenter's action.

Despite the compelling nature of Maier's account, we nevertheless note that the interpretation of his findings is rendered inconclusive because of a failure to use a control condition in which participants received no hint. Without such a control condition it is unclear whether participants benefited simply from having increased exposure to the problem (see Landrum 1990, for a failure to find a hint effect when using a no-hint control group as a comparison condition). Other research, however, has provided convincing evidence that explicit hints operating at a conscious level can facilitate insight into the two-string problem. For example, Battersby et al. (1953) found that reduced solution latencies on the two-string problem could be affected by simply highlighting objects within the room that might be of relevance to a solution.

Results from a recent study by Thomas and Lleras (2009b) have provided some of the most compelling evidence for the role of *non-conscious* hints in guiding insight in the two-string problem. In their study, Thomas and Lleras asked participants to attempt the problem while occasionally taking exercise breaks during which they moved their arms either in a manner related to the problem's solution (the "swing" group) or in a manner inconsistent with the solution (the "stretch" group).

Although a majority of participants were unaware of the relationship between their arm-movement exercises and the problem solving task, individuals in the swing group were significantly more likely to solve the problem than individuals in the stretch group.

Thomas and Lleras claim that their findings are consistent with theories of "embodied cognition" (e.g., Barsalou 1999; Gibbs 2006; Spivey 2007; Wilson 2002; Zwaan 1999), which propose that to understand how the mind accomplishes its goals (e.g., solving problems) one has to understand the mind in the context of its links to a *body* that interacts with the physical world. In this way, embodied cognition theories propose that knowledge representations in the brain maintain properties of the sensorimotor states that gave rise to them in the first place (e.g., Barsalou 1999). More recently, Barsalou (2008) has defended a more general account of "grounded cognition", which proposes that much cognition is underpinned by modal simulations (including mental imagery), bodily states and situated actions. Hutchins (2010) evokes the related concept of "enactment" to argue how enacted multimodal representations are involved in the construction of memories for the past, the experience of the present, and the attainment of future goals, as in problem solving.

Thomas and Lleras's (2009b) results with the two-string problem fall neatly within these embodied, enacted and grounded views of cognition, since their findings show how bodily actions can influence thinking, such that people can be guided implicitly toward insightful solutions by directing their actions in solution-appropriate ways. Over the past decade there have, in fact, been a wealth of "demonstration experiments" (Barsalou 2010) within the general area of grounded cognition, which show how grounded mechanism seem to lie at the heart of goal-directed cognition that is related to perception and action, memory, conceptual processing, language comprehension, social judgement and thought (see Barsalou 2010, for a highly focused review of this compelling evidence).

The embodied cognition effects in insight problem solving that have been identified by Thomas and Lleras (2009b) clearly take the concept of "hints" in an important new direction, since what is being evidenced appears to be the implicit perceptual priming of motor representations in the brain that are associated with "swinging" movements. We still view such embodied priming effects as arising from a form of *hint*, but a hint that is nevertheless a long way removed from a direct verbal instruction, an explicit gestural prompt or an implicit environmental cue. Our own particular interest is in yet another type of external hint, which is of the kind that can derive from other people's eye movements, where such eye movements can serve to convey information and direct attention and may also provide a basis for imitative motor simulations.

Hints arising from following other people's eye movements are clearly distinct from normal language-based or gesture-based directives, being both more subtle and highly multifaceted in nature. We review the way in which eye movements can act as hints in the next section, where we examine the communicative properties of other people's eye movements, such as their capacity to convey information and to direct our attention (e.g., Kleinke 1986; Sheperd 2010), and where we also entertain the role of eye movements in priming task-relevant perceptual and motor simulations in an observer. Suffice it to say for now that eye movements seem to engender a wide array of cueing effects, ranging from implicit and embodied priming of motor sequences and perceptual simulations right through to explicit and directive communicative prompts.

Hints Arising from Other People's Eye Movements

Many studies have demonstrated that we are highly sensitive to other people's gaze (Gibson and Pick 1963; Symons et al. 2004) and that this sensitivity develops very early in life, with even young infants showing sophisticated gaze following behaviour (Brooks and Meltzoff 2005; Corkum and Moore 1995). Following someone else's gaze can guide attention towards a particular object or item, which in turn enables knowledge to be inferred or to be directly shared, as is the case when observers verbalise their mutual interest in a jointly attended item (Flom and Pick 2007; Hanna and Brennan 2007). Such "joint attention" has been claimed to form a basis for much early learning that arises in development (Butterworth and Jarret 1991), although it also remains ubiquitous in adulthood (Driver et al. 1999). Adult observers are particularly adroit at taking into account nearby objects when following another's gaze, with neuropsychological evidence indicating that following a person's gaze can function to transfer the *intentionality* of that person to the observer (Becchio et al. 2008; Lobmaier et al. 2006). Therefore, by watching where another person looks we alter our own processing of objects. At a very minimum, we obtain cues from where the other person is attending that have the potential to be highly beneficial for our own cognitive activities—such as problem solving—especially if we are in the presence of an individual who has greater domain expertise than we do.

Given the ubiquity of attention-directing eye movements that arise constantly in our everyday work and social environments it would seem that a great deal of ongoing high-level cognition in areas such as problem solving, reasoning, judgement and decision making might be shaped by such external cues. Several recent studies have examined this issue by investigating the value of the direct presentation of the eye movement patterns of one observer to another observer. The aim of these studies is to assess how this form of attentional guidance can improve performance in a range of visual search and problem solving tasks. Experts are known to look at task-relevant areas more often than novices and also demonstrate more effective search strategies (e.g., Chapman and Underwood 1998; Charness et al. 2001; Krupinski 1996), which means that there is a growing belief that an expert's eye movement patterns should be particularly useful in training novices where to look (Gegenfurtner et al. 2011; Jarodzka et al. 2012; Krupinski et al. 2006; Nalanagula et al. 2006).

By recording the eye movement behaviour of experts and showing the projected fixation position to other observers, recent studies have demonstrated that novices can detect more faults during aircraft fuselage inspection (Sadasivian et al. 2005) and during circuitry board inspection (Nalanagula et al. 2006). In addition, novices can benefit from such cues so as to make better clinical decisions (Jarodzka et al. 2012; Litchfield et al. 2010). For tasks such as problem solving and program debugging, which involve factors in addition to visual search, viewing another's gaze can result in shorter task completion times (Pomplun et al. 1996; Stein and Brennan 2004; Velichkovsky 1995). Of course, there are some situations where unambiguous verbal comments are generally preferred to another's gaze (e.g., Bard et al. 2007; Van Gog et al. 2009). As such, positive effects will likely depend on the task demands and the different ways in which eye movement patterns can be presented (cf. Jarodzka et al. 2012; Nalanagula et al. 2006). We return to some of these issues in a subsequent section below where we report some recent research that we conducted to examine these questions.

A final issue of importance concerns the fact that projecting gaze behaviour to observers involves "artificially" represented gaze (i.e., eye-movement patterns dynamically overlaid on a screen-based image of a task or problem), and therefore it is debatable whether the cognitive processes that are evoked are the same as those arising during normal gaze perception and gaze following. For example, the emotional expression of the person being watched is often taken into account when using normal gaze as a predictive cue (Bayliss et al. 2007) or when determining mutual gaze (Lobmaier et al. 2008). Although this information is absent when using artificial gaze, this method does provide an opportunity to look at more complex gaze sequences rather than simple directional processes. Indeed, there have been reports of higher-order search strategies based on another's real-time gaze. For example, observers can regulate their own search behaviour in a collaborative visual search task by strategically ignoring areas that they can see are being observed by their collaborator (Brennan et al. 2008). By taking advantage of these non-verbal gaze cues, observers have been shown to be able to reduce their search times significantly. Thus, even with artificially represented gaze, observers alter their behaviour based on the perceived processing of others.

Making Use of Attentional Hints and Gaze Cues in Insight Problem Solving

As outlined above, "insight" in problem solving arises when an individual who is stuck on a task and unable to make headway suddenly breaks free of their unhelpful thoughts and is able to find a solution. The two-string problem (Maier 1931), introduced earlier, is a classic example of an insight problem; another example is Duncker's (1945) radiation problem, where participants have to find a way to destroy a stomach tumour using lasers without harming surrounding healthy tissue (see Fig. 12.1a). Although the solution involves only two critical components (i.e., converging multiple lasers of low intensity), problem solvers typically reach an impasse, and few solve the problem without hints such as visual analogies (Pedone et al. 2001). Resolving the impasse requires some form of restructuring, with theories typically emphasising the role of unconscious constraint relaxation (Knoblich et al. 1999; Ohlsson 1992).

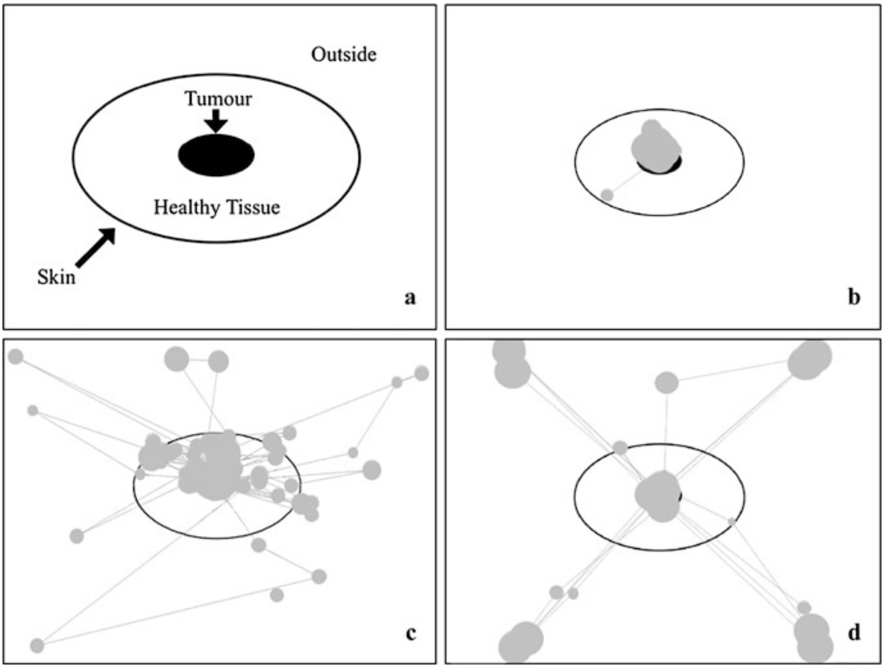


Fig. 12.1 Panel (**a**) shows a diagrammatic representation of Duncker's (1945) radiation problem, as used by Litchfield and Ball (2011), which was accompanied by a statement of the problem as follows: "Given a patient with an inoperable stomach tumour, and lasers which destroy organic tissue at sufficiently high intensity, how can one cure the patient with these lasers and, at the same time, avoid harming the healthy tissue that surrounds the tumour?" Panels (**b**), (**c**) and (**d**) each depict a static scanpath that represents the 30 seconds of dynamic eye movement behaviour that was shown to participants in Litchfield and Ball's tumour-fixation condition (**b**), in their natural skin-crossing condition (**c**), and in their didactic skin-crossing condition (**d**). The dynamic scanpath that was presented to participants took the form of a moving blue gaze cursor

Recent research has examined the way in which eye movements and attentional guidance might influence solution success. Grant and Spivey (2003) demonstrated that participants who solve Duncker's (1945) radiation problem have specific fixation distributions just prior to solving the problem that are localised between the skin areas surrounding the tumour. They hypothesised that these fixation patterns may reflect the solver's mental imagery of where the lasers would have to be fired from to destroy the tumour. To test this hypothesis, Grant and Spivey conducted an experiment in which they made the diagram of the skin area more conspicuous by subtly increasing and decreasing its size. This "animated skin" condition engendered double the rate of solution success as seen in the control conditions, suggesting that implicitly guiding participants' attention to the skin areas primed the perceptual-motor pattern associated with the solution. As they put it, "in-and-out eye movements themselves may have served as an embodied physical mechanism that jump-started a perceptual simulation (Barsalou 1999) of multiple incident rays, and wound up supporting the inference that multiple lasers could be fired (at low intensities) from different points outside the diagram" (p. 466).

Thomas and Lleras (2007) examined this perceptual-motor relationship further by sporadically guiding participants' eye movements (via an unrelated number tracking task), either in a pattern related to the solution to Duncker's radiation problem or in various unrelated patterns. Importantly, those participants who moved their eyes in a pattern related to the problem's solution were more likely to solve the problem. Thomas and Lleras (2007) proposed that the improvement in solution rates arose neither because these cues increased the salience of the outer areas, nor because they increased the number of skin-crossing saccades. Instead, the improvement was a consequence of evoking a specific perceptual-motor pattern that embodied the solution to the problem, with the underlying shift of attention that accompanied the pattern being the crucial factor, rather than volitional eye movement per se. This interpretation was supported by Thomas and Lleras (2009a) in their follow-up research, which used the same paradigm as their earlier study (i.e., involving an unrelated number tracking task), but with participants this time having to do the tracking by shifting only their *attention* while keeping their eyes fixed on the centre of the display. This "attention-shift" condition produced similar solution facilitation to that which arose in the condition that permitted free eye movements during the tracking manipulation. Collectively, these results indicate that the attentional shifts that arise during scene examination can act as valuable, implicit hints that can guide how people think and reason.

The research that we have reviewed so far in this section demonstrates how eye movements and attentional shifts that derive from the *problem solver* can function to bootstrap their own efforts at insightful problem solving. But what about the situation where the problem solver has access to the eye movement patterns of another solver? We have recently examined this question—also in the context of participants tackling Duncker's radiation problem (see Litchfield and Ball 2011). For this problem, it would be impractical to use a face-to-face model to convey the sequence of solution-specific eye movements described by Thomas and Lleras (2007). Instead, we decided to show the previously recorded eye movement patterns of a successful solver to observers so as to indicate where that individual looked during the task. As discussed earlier, although viewing "artificially" represented gaze is unlikely to involve identical cognitive processing to that which is deployed during normal gaze-following behaviour, viewing another's eye movement patterns still allows observers to modify their information processing contingent upon the processing of others (e.g., Brennan et al. 2008; Nalanagula et al. 2006; Neider et al. 2010; Stein and Brennan 2004; Velichkovsky 1995).

A further issue addressed in our study concerned whether the model providing the eye movements was actually aware that other observers might be using their eye movements as visual cues. Gaze behaviour is often interactively "regulated" when people know that their eye movements are being observed during face-to-face situations (Kleinke 1986) or that their eye movements will be projected to others via eye-tracking equipment (Neider et al. 2010; Velichkovsky 1995). Indeed, the model providing the eye movements can deliberately *control* their gaze so that their intent is more effectively communicated to observers. There may, therefore, be a difference between viewing the eye movement patterns of a successful problem solver who is unaware that their eye movements are to be used as subsequent cues versus viewing another's eye movements as they actively try to convey the solution of the problem in a didactic manner (Velichkovsky 1995).

In our study (Litchfield and Ball 2011) we examined problem solving in three conditions (see Fig. 12.1). These conditions were: (1) "tumour-fixation", where participants viewed a problem solver's eye movement patterns focusing solely on the central tumour; (2) "natural skin-crossing", where participants viewed a problem solver's eye movement patterns *naturally* making skin-crossing saccades between the outside area and the central tumour from multiple directions; and (3) "didactic skin-crossing", where participants viewed eye movement patterns that *deliberately* made skin-crossing eye movements between the outside area and the central tumour from multiple directions. Performance in the tumour-fixation condition acted as a control, since this condition would not cue participants to make skin-crossing saccades associated with the solution.

As predicted, our results showed that participants who were encouraged to make skin-crossing saccades were more likely to solve the problem than those who simply focused on the tumour. Although the natural and didactic conditions led to equivalent final solution rates, participants in the didactic condition showed reduced solution latencies (i.e., faster insight solutions) relative to those in the natural condition. Overall, Litchfield and Ball (2011) conclude that there is good evidence to support the conclusion that viewing where another person looks can guide attention in a way that can positively affect problem solving.

Extending the Use of Gaze Cues to Real-World Problem Solving

One of the clear advantages of cueing visual attention by means of task-specific eye movement hints is the general applicability of this method to a wide range of real-world tasks. For example, in addition to studying how another's gaze can provide a problem solver with insight into how to treat a hypothetical tumour, we have also examined how the use of eye movement cues might facilitate problem solving performance in the domain of diagnostic radiography. Our particular research focus in this domain was on diagnosticians identifying pulmonary nodules (pre-cursors to lung cancer) in chest x-ray inspection (Litchfield et al. 2010). The eye movement patterns shown in this study were not designed to simulate a movement-based *thought process* relating to the solution to the problem (as in Litchfield and Ball 2011), but were instead presented so as to encourage observers to undergo the same series of *bodily movements* (i.e., eye movements) that experts make when examining chest x-rays and to focus on the same regions of interest that attract the attention of experts. In this way, these non-verbal, step-by-step hints served to influence observers' evaluative decisions regarding the presence of pulmonary nodules, and by doing so guided thought indirectly by first guiding attention.

In our study we used signal detection theory to assess diagnostic performance. In our initial experiment we found that both novice and experienced observers performed better at detecting lung nodules in an x-ray when previously shown the search behaviour of either a novice radiographer or an expert radiologist viewing the same visual display. In follow-up experiments we manipulated the task specificity of the presented gaze patterns (i.e., whether or not they related to lung nodule detection) as well as the perceived expertise of the model providing the gaze cues. These subsequent experiments established that only novices consistently improved their performance when shown an expert's search behaviour, and that these benefits only arose when the eye movements shown were related to a task-specific search. Moreover, a detailed examination of the contribution of model expertise indicated that even a naïve observer's search behaviour could help scaffold the decisions made by novices. Such evidence provides further support for the general applicability of gaze following behaviour and the pervasive functions associated with interpreting another's gaze (Emery 2000; Tomasello 1999). Our collective findings highlight that gaze following is a ubiquitous behaviour that can be used to support a variety of decisions and judgements in problem solving contexts. It is clear that knowing where another person looks can cue attention and shape thoughts and decisions, even when the problem solver is not in a face-to-face situation.

Conclusion

Over the past decade the theoretical conception of problem solving as being situated, embodied and interactive in nature has gained considerable momentum. One aspect of this changing view of problem solving has been the appreciation that as a highly social species, whenever humans are faced with a difficult task the most natural tendency is for them to exploit all externally available *hints* that provide clues as to how to proceed. Such hints may derive from looking (literally) to others for advice or from making use of other prompts that are available within the social and physical environments (Kirsh 2009; Kleinke 1986; Sheperd 2010; Tomasello 1999).

Our discussion in this chapter has taken much inspiration from Kirsh's (2009) recent sketch of a "theory of hints", which he presented as a first step toward a detailed consideration of what hints entail, where they stem from and how they provide a key basis for understanding how people overcome problems in concrete, real-world settings through a dynamic interaction with all elements of their extended environment. In contributing to Kirsh's theory of hints we have attempted to broaden the conception of hints to include the communicative cues that derive from other people's eye movements that can serve to direct attention and impart information in ways that can facilitate reasoning and problem solving. In reviewing a wide range of evidence for the role of eye movement cues as scaffolds for people's problem solving we have touched upon key notions in the emerging literature on grounded cognition (e.g., Barsalou 2008) and embodied cognition (e.g., Gibbs 2006; Spivey 2007).

There is no doubt that the role of eye movement cues in facilitating problem solving is complex and multi-faceted. Effects that have been observed include the embodied priming of motor sequences and perceptual simulations that enable problem solvers to enact the thought processes that underpin the solution to a problem (e.g., Litchfield and Ball 2011). Other effects involve implicitly cueing observers of eye movement patterns to imitate similar eye movement patterns when searching visual displays for possible pathology (e.g., Litchfield et al. 2010). This latter form of cueing brings with it a key *attentional* component in as much as imitative eye movements guide attention to regions of interest within the problem display, which, in turn, become subjected to inferential thought aimed at making a diagnostic decision. Eye movements can also function in many communicative contexts as an *explicit* cue to archive "joint attention", ensuring, for example, that people are attending to the same objects within the immediate environment and can thereby engage in shared reasoning about such objects (e.g., Becchio et al. 2008; Flom and Pick 2007; Hanna and Brennan 2007).

We conclude by suggesting that if cognitive science is serious about considering intelligence as an interaction between individuals and the hints and cues that derive from the physical and social environments (e.g., Barsalou 2010; Kingstone et al. 2003; Proffitt 2006; Sebanz et al. 2006; Wilson 2002), then considerable research time and effort will need to be devoted to developing the theories and methodologies outlined in this chapter. It is only by doing so that we will truly be able to quantify complex problem solving and reasoning behaviour from a situated, embodied and grounded perspective.

Acknowledgements This chapter is based on a paper presented at the Symposium on Distributed Cognition, University of Kingston, Kingston-upon-Thames, UK, held in July 2010, and sponsored and organised by Kingston Psychology and the Distributed Language Group. We are grateful for David Kirsh's invaluable feedback on this presentation, which inspired the connection between his sketch of a theory of hints and our interest in embodied and interactive eye movement cues in problem solving. We are also grateful to Stephen Cowley and Frédéric Vallée-Tourangeau for follow-up discussions on the work we presented at the symposium as well as for their critical comments on the present chapter.

References

Ball, L. J., & Ormerod, T. C. (2000a). Applying ethnography in the analysis and support of expertise in engineering design. *Design Studies*, *21*, 403–421.

Ball, L. J., & Ormerod, T. C. (2000b). Putting ethnography to work: the case for a cognitive ethnography of design. *International Journal of Human-Computer Studies*, *53*, 147–168.

Bard, E. G., Anderson, A. H., Chen, Y., Nicholson, H. B. M., Havard, C., & Dalzel-Job, S. (2007). Let's you do that: sharing the cognitive burdens of dialogue. *Journal of Memory and Language*, *57*, 616–641.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Barsalou, L. W. (2010). Grounded cognition: past, present, and future. *Topics in Cognitive Science*, *2*, 716–724.

Battersby, W. S., Teuber, H. L., & Bender, M. B. (1953). Problem solving behavior in men with frontal and occipital brain injuries. *The Journal of Psychology*, *35*, 329–351.

Bayliss, A. P., Frischen, A., Fenske, M. J., & Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition*, *104*, 644–653.

Becchio, C., Bertone, C., & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences*, *12*, 254–258.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: the costs and benefits of shared gaze during collaborative search. *Cognition*, *106*, 1465–1477.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*, 535–543.

Busby, J. S. (2001). Practices in design concept selection as distributed cognition. *Cognition Technology & Work*, *3*, 150–160.

Butterworth, G., & Jarret, N. (1991). What minds have in common in space: spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, *9*, 55–72.

Chapman, P. R., & Underwood, G. (1998). Visual search of driving situations: danger and experience. *Perception*, *27*, 951–964.

Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: evidence from eye movements. *Memory & Cognition*, *29*, 1146–1152.

Corkum, V., & Moore, C. (1995). Development of joint visual attention in infants. In C. Moore & P. J. Dunham (Eds.), *Joint attention: its origins and role in development* (pp. 61–83). Hove: Erlbaum.

Dominowski, R., & Dallob, P. (1995). Insight and problem solving. In R. J. Sternberg & J. Davidson (Eds.), *The nature of insight* (pp. 33–62). Cambridge: MIT Press.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, *6*, 509–540.

Dunbar, K., & Blanchette, I. (2001). The in vivo/in vitro approach to cognition: the case of analogy. *Trends in Cognitive Sciences*, *5*, 334–339.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, *58*(5 (270)), i–113. doi:10.1037/h0093599.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, *24*, 581–604.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.

Flom, R., & Pick, A. D. (2007). Increasing specificity and the development of joint visual attention. In R. Flom, K. Lee, & D. Muir (Eds.), *Gaze-following: its development and significance* (pp. 95–111). London: Erlbaum.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*, 523–552.

Gibbs, R. (2006). *Embodiment in cognitive science*. New York: Cambridge University Press.

Gibson, J., & Pick, A. (1963). Perception of another person's looking behavior. *The American Journal of Psychology*, *76*, 386–394.

Gilhooly, K. J., & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking & Reasoning*, *11*, 279–302.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: guiding attention guides thought. *Psychological Science*, *14*, 462–466.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*, 596–615.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.

Hutchins, E. (2010). Enaction, imagination, and insight. In J. Stewart, O. Gapenne, & E. A. Di Paolo (Eds.), *Enaction: toward a new paradigm for cognitive science*. Cambridge: MIT Press.

Jarodzka, H., Baslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., & Eika, B. (2012). Conveying clinical reasoning based on visual observation via eye movement modeling examples. *Instructional Science*, *40*, 813–827.

Kaplan, C., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, *22*, 374–419.

Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., & Eastwood, J. D. (2003). Attention researchers! It's time to take a look at the real world. *Current Directions in Psychological Science*, *12*, 176–184.

Kirsh, D. (2009). Problem solving and situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition*. Cambridge: Cambridge University Press.

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychological Bulletin*, *100*, 78–100.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*, 1534–1555.

Kokinov, B., Hadjiilieva, K., & Yoveva, M. (1997). Is a hint always useful? In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual conference of the cognitive science society*. Hillsdale: Erlbaum.

Krupinski, E. A. (1996). Visual scanning patterns of radiologists searching mammograms. *Academic Radiology*, *3*, 137–144.

Krupinski, E. A., Tillack, A. A., Richter, L., Henderson, J. T., Bhattacharyya, A. K., Scott, K. M., et al. (2006). Eye-movement study and human performance using telepathology virtual slides: implications for medical education and differences with experience. *Human Pathology*, *37*, 1543–1556.

Landrum, R. E. (1990). Maier's two-string problem revisited: evidence for spontaneous transfer. *Psychological Reports*, *67*, 1079–1088.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth a thousand words. *Cognitive Science*, *11*, 65–99.

Lave, J. (1988). *Cognition in practice: mind, mathematics and culture in everyday life*. Cambridge: Cambridge University Press.

Litchfield, D., & Ball, L. J. (2011). Using another's gaze as an explicit aid to insight problem solving. *Quarterly Journal of Experimental Psychology*, *64*, 649–656.

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology. Applied*, *16*, 251–262.

Lobmaier, J. S., Fischer, M. H., & Schwaninger, A. (2006). Objects capture perceived gaze direction. *Experimental Psychology*, *53*, 117–122.

Lobmaier, J. S., Tiddeman, B. P., & Perret, D. I. (2008). Emotional expression modulates perceived gaze direction. *Emotion*, *8*, 573–577.

Maier, N. R. F. (1931). Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *Journal of Comparative & Physiological Psychology*, *12*, 181–194.

Nalanagula, D., Greenstein, J. S., & Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics*, *36*, 289–300.

Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., & Zelinsky, G. J. (2010). Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review*, *17*, 718–724.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Ormerod, T. C., & Ball, L. J. (2007). Qualitative methods in cognitive psychology. In C. Willig & W. Stainton-Rogers (Eds.), *Handbook of qualitative research in psychology* (pp. 553–574). London: Sage.

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. T. Keane & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (Vol. 1, pp. 1–44). London: Harverster Wheatsheaf.

Pedone, R., Hummel, J. E., & Holyoak, K. J. (2001). The use of diagrams in analogical problem solving. *Memory & Cognition*, *29*, 214–221.

Pomplun, M., Ritter, H., & Velichkovsky, B. (1996). Disambiguating complex visual information: towards communication of personal views of a scene. *Perception*, *25*, 931–948.

Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science*, *1*, 110–112.

Reid, F. J. M., Culverhouse, P. F., Jagodzinski, A. P., Parsons, R., & Burningham, C. (2000). The management of electronics engineering design teams: linking tactics to changing conditions. *Design Studies*, *21*, 75–97.

Sadasivian, D., Greenstein, J. S., Gramopadhye, A. K., & Duchowski, A. T. (2005). Use of eye movements as feedforward training for a synthetic aircraft inspection task. In *Proceedings of the SIGCHI conference on human factors in computing systems, CHI 2005* (pp. 289–300). New York: ACM Press.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*, 70–76.

Sheperd, S. V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience*, *4*, 1–13.

Simon, H. A. (1981). *The sciences of the artificial* (2nd ed.). Cambridge: MIT Press.

Smith, R. W., & Kounios, J. (1996). Sudden insight: all or none processing revealed by speed-accuracy decomposition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *22*, 1443–1462.

Spivey, M. (2007). *The continuity of mind*. Oxford: Oxford University Press.

Stein, R., & Brennan, S. E. (2004). Another person's eye gaze as a cue in solving programming problems. In *Proceedings of ICMI 2004: sixth international conference on multimodal interfaces* (pp. 9–15). New York: ACM Press.

Suchman, L. (1987). *Plans and situated actions: the problem of human-machine communication*. Cambridge: Cambridge University Press.

Symons, L. A., Lee, K., Cedrone, C. C., & Nishimura, M. (2004). What are you looking at? Acuity for triadic eye gaze. *The Journal of General Psychology*, *131*, 451–469.

Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: a computational model of multiple representations. *Cognitive Science*, *21*, 305–350.

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: on the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review*, *14*, 663–668.

Thomas, L. E., & Lleras, A. (2009a). Covert shifts of attention function as an implicit aid to insight. *Cognition*, *111*, 168–174.

Thomas, L. E., & Lleras, A. (2009b). Swinging into thought: directed movement guides insight in problem solving. *Psychonomic Bulletin & Review*, *16*, 719–723.

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard: Harvard University Press.

Van De Weil, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (2000). Knowledge restructuring in expertise development: evidence from pathophysiological representations of clinical cases by students and physicians. *European Journal of Cognitive Psychology*, *12*, 323–356.

Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, *25*, 785–791.

Velichkovsky, B. M. (1995). Communicating attention: gaze position transfer in cooperative problem solving. *Pragmatics & Cognition*, *3*, 199–222.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625–636.

Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*, 81–88.

Chapter 13
Naturalising Problem Solving

Frédéric Vallée-Tourangeau and Gaëlle Villejoubert

Abstract A striking feature of people engaged in problem solving outside the psychologist's laboratory is that it exhibits a great deal of interactivity in a physical space populated with external symbols, artefacts, and, of course, other people. Yet, problem-solving researchers often design experimental procedures in which interactivity is either limited or eliminated. We review traditional areas of problem solving research and introduce new experimental methodologies wherein problems can only be solved by manipulating or restructuring a physical space. In all instances, problem-solving performance is markedly superior than when observed in two-dimensional non-interactive contexts. We suggest that the nature of the processes engaged in solving problems in distributed environments is different than in static environments and should encourage cognitive psychologists to revisit the process models elaborated to account for problem solving behaviour traditionally recorded on the basis of an unmodifiable problem presentation.

In adapting to their environment, humans[1] solve problems. This activity takes place among a vast and dynamic array of artefacts: "layers of artefacts saturate almost everywhere we go" (Kirsh 2009, p. 270). To be sure, artefacts such as calculators, data management software, computers can facilitate complex computations. But others, of more modest complexity, such as the pen and paper, can help articulate and structure thinking. And once written down (on the back of an envelope, a notebook) these thoughts acquire permanency and mobility (Latour 1986; Kirsh 2006, 2010) and can act as artefacts to support and guide further thinking. Artefacts enhance both prospective memory and retrospective memory, as well as augment working memory (Vallée-Tourangeau 2013).

[1]There is a vast literature on how non-human organisms solve problems that will not be reviewed here (see for example Pearce 2008).

F. Vallée-Tourangeau (✉) · G. Villejoubert
Department of Psychology, Kingston University, Kingston upon Thames, Surrey, KT1 2EE, UK
e-mail: f.vallee-tourangeau@kingston.ac.uk

G. Villejoubert
e-mail: g.villejoubert@kingston.ac.uk

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_13,

In addition, people naturally seek to change the physical environment to make thinking easier and more efficient (Kirsh 1996). Space itself is a resource that can facilitate thinking, that is it can be structured, designed (and redesigned) to create a cognitively congenial environment (Kirsh 1995). Solving jigsaw puzzles involves physically juxtaposing different pieces to gauge their fit; in Scrabble, letter tiles are physically rearranged to facilitate word production; in Tetris, tetrominoes are physically rotated to determine their optimal place along a line. And beyond puzzles and games, experts structure an external environment to support fast and effortless thinking (Kirsh 1995). Scientists use physical objects and their arrangement in space to think—Watson (1968, pp. 123–125) describes how he cleared his desk, cut out shapes corresponding to the four atomic bases, and manipulated them until he saw which ones could be paired to hold the double helix together.

Artefacts recruited in thinking are rich, varied and modifiable. Their recruitment is at times strategic, such that their users actively engage in their design and engineer their function, and at others, opportunistic, that is they are picked up from the environment in an ad hoc fashion to help solve a problem (capitalizing on a fortuitous interaction). The modification of the physical space may be guided by well-formed hypotheses, but at other times the modifications may be less strategic. Still, as we will illustrate below, such changes inevitably transform the affordance landscape which may lead to important changes in the representation of the problem, in turn enabling thinkers to identify new paths to solution.

The psychology of problem solving largely ignores the role of these artefacts and these interactive and space-reshaping strategies as part of the process of thinking. This choice is driven by a number of interrelated theoretical and methodological reasons. The commonly endorsed methodological exigencies of cognitive psychology research impose a strict control of the experimental environment fashioned to examine thinking. As a result there are strong constraints that govern the selection of tasks and problems used to observe problem solving behaviour. These problems tend to be well defined and are presented in an experimental setting in which participants can rarely effect changes, either on the basis of instructional guidance and or because the problem is presented on a piece of paper or a computer screen which simply does not offer the reasoner the possibility to manipulate a physical presentation of the problem. Participants have little opportunity to exploit naturally occurring affordances from concrete objects that contribute to the presentation of the problem or to reshape the epistemic landscape of their environment by restructuring it. A case in point: Luchins's celebrated demonstration of mental set using a series of water measurement problems (the water jar task). In this task participants learn to solve a series of transformation problems by applying a moderately complex pouring maneuver, which once discovered can be applied to all remaining problems to yield the correct solution. However, for some later problems, a simpler maneuver can be employed, but participants persevere in using the relatively more complicated one, to the point of even failing to solve a very simple problem for failing to 'see' how a simpler set of transformation can lead to the right answer. Yet, this water jar task involves neither jars nor the manipulation of any actual liquid: It is a pen and paper task (e.g., Luchins 1942, and many subsequent replications, e.g., McKelvie 1990). To be sure, pen and paper also configure an extended cognitive system, but a system that offers a narrower set of perceptual experiences, primes certain arithmetic schemas—that cues ways of solving arithmetic puzzles in formal teaching environments—and affords a limited repertoire of actions, that substantially constrain how an otherwise physically embedded agent can reconfigure the space and the elements that compose it to arrive at an efficient problem solution.

Process models of transformation problems such as river crossing problems are based on data obtained either using pen and paper tasks or simple computer interfaces that represent schematised riverbanks and objects designed to record state changes, not to permit manipulation (e.g., Jeffries et al. 1977). Knoblich et al's (1999) important paper on constraint relaxation and chunk decomposition using matchstick algebra problems employs an experimental procedure where the matchstick problems are presented on a computer screen; participants can not manipulate them. It is also of interest to note that in all those examples the tasks refer to familiar and concrete artefacts—matches, water jars, animals—and yet those are never presented as concrete manipulable artefacts. This may reflect the presumption that it is the same thing to think in the head as it is to think with objects in the world. In contrast, developmental psychologists who worked with the river crossing task, being less sanguine about 'formal operations' presumably, have taken care to design rich interactive thinking environments with physical materials representing the boat, the river, and figurines corresponding to the cover story characters (e.g., Gholson et al. 1987; see also Vallée-Tourangeau et al. 2013).

The development of computational models of search in the 1960s (e.g., Simon and Newell 1962) drove much of the research on problem solving and conspired to entrench even more strongly certain theoretical commitments. These models of thinking implicate a mental representation of a problem space and sequential state transformation through that space using well defined rules and operators. In the elaboration of these models, people and artefact are decoupled. Yet, with our ethnographer's hat, observing people thinking in quotidian (Lave 1988) or more formal scientific environments (e.g., Giere and Moffatt 2003) we are struck by the ubiquitous interactions with artefacts. The scientific method that the psychologist brings to bear on problem solving however extricates people from this rich set of interactions and disembodies the mind (Hutchins 1995). We would thus argue that the perspective on problem solving offered by traditional cognitive psychology is profoundly unrepresentative of thinking as it occurs outside the laboratory.

The remainder of this chapter reviews in greater details two important areas of research that have informed and defined the psychology of problem solving in the past 70 years. The nature of the experimental methodology in those areas is critically examined. We then review recent data that were obtained with a new research methodology that couches these problems in physical environments that afford mutability and change. In these reasoning contexts, participants interact with the physical constituents of the problem presentation. These interactive problem-solving environments yield performance that departs significantly from what is commonly observed and traditionally recorded. These new data inform the development of different process models of problem solving behaviour, in terms of distributed representation and

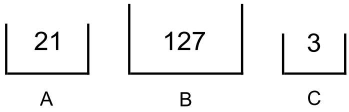


Fig. 13.1 A volume measurement problem as used in Luchins (1942). The goal is to obtain exactly 100 units of volume. In this case, this can be achieved by filling B, from B filling A once, and then from B filling C twice, or B − A − 2C

control. The more general implications of these findings for the development of the psychology of problem solving will be discussed in a final section.

The Mechanization of Thought: Einstellung

The water jar task introduced by Luchins (1942) is offered as a window into perseverance in problem solving driven by previous experience (a more recent treatment of mental set is offered in Bilalić et al. 2008, 2010). The water jar task is a series of volume measurement problems designed to showcase how experience and past practice may 'mechanize' thinking, infusing it with a rigidity that prevents problem solvers from discovering simpler, more efficient means to solve problems. In the procedure designed by Luchins, participants are presented with a series of simple arithmetic problems with 3 water jars (labeled A, B and C) for each of which the goal is to obtain an exact volume of liquid by filling and pouring water from one jar into another in a certain order until one of the jars contains the target amount. Figure 13.1 illustrates a typical problem. Problems are presented in two blocks, although participants are not informed of this ordering. The first block is composed of five problems for which the solution can always be derived with the rule B − A − 2C (as in Fig. 13.1). These problems are referred to as *einstellung* or *set* problems. In the second block, there are four *critical* problems and one *extinction* problem. All can be solved using a much simpler rule, either A − C or A + C. All critical problems can also be solved using the more complicated rule B − A − 2C, while the extinction problem can only be solved with A − C.

Participants are split in two groups, those who are presented the set problems first, before solving the critical problems, and those who start with the critical problems first, without being exposed to the set problems. These latter participants show little difficulty in discovering the simple rules such as A − C or A + C to solve the critical problems—and the majority can solve the extinction problem that can only be solved with the rule A − C. However, the participants who trained on the set problems and discovered the B − A − 2C rule to solve them, persevere with the more complicated solution, rarely employing the simpler rules. More important, the majority of these participants fail to solve the extinction problem that can only be solved with the simpler rule. The Water Jar set effect is an important demonstration of how past experience can shape problem solving strategies such that simpler, more efficient, and less costly rules to solution are overlooked. The interventionist challenge for psychology and education is how to enhance creativity and foster open-minded thinking such as to reduce the mechanization of thought on the basis of prior experience.

Numerous replications of the phenomenon since Luchins (1942) published his monograph and its theoretical importance have overshadowed one extraordinary aspect of the experimental procedure: Participants in these experiments are not asked to manipulate actual water jars in trying to solve the problems.[2] This is a remarkable feature of the experimental procedure: Participants do not interact with a physical environment to solve these problems. The calculations are simulated mentally, the various pouring maneuvers abstracted as mental number manipulations. Hence, the 'water jar' problem is a mental puzzle, a brain teaser, designed for disembodied and decontextualised minds. In contrast, people solve problems with hand and body in a world flush with affordances.

Presenting the water jar problem in a physical environment may augment the range of processes that are engaged in solving the task by adding action and vision to imagery and projection (Kirsh 2009). A richer environment could provide a broader set of clues to guide and constraint action—behavioural control could be distributed in a more complex manner across internal and external cues. This enrichment of resources and processes could prevent or attenuate the mechanization of thought. Hence we deemed it important to revisit this classic demonstration of the mechanization of thought in an environment that offered rich perceptual cues and a range of actions.

We replicated Luchins's (1942) experiment using the 11 original problems (1 practice +10 test problems). Among the 10 test problems, the first five were the set problems for which the rule $B - A - 2C$ was the only one that yielded the solution (Vallée-Tourangeau et al. 2011). For the remaining five problems, the acquired rule yielded a correct solution for four of them, but a simpler $A + C$ or $A - C$ rule provided a correct solution for all of them. Our participants were undergraduate students who were allocated to one of two groups. Participants in the first group were presented with the 10 test problems as a pen and paper task. Participants in the second group were invited to solve the water jar problems by interacting with sets of three jars at a sink: They filled jars and poured water from one to the other to obtain a target amount. The fourth test problem (a set problem) as presented in both groups is illustrated in Fig. 13.2.

Participants in both groups learned the $B - A - 2C$ rule easily: Mean percent use of the rule is nearly at ceiling for both groups (see left panel of Fig. 13.3). The critical data is the degree of perseverance or 'mechanization' of thought during the second block of problems (see middle panel of Fig. 13.3). Participants in the interactive group were much less likely to persevere using the more complicated B

[2] In a paper that summarized their efforts to reduce perseverance Luchins and Luchins (1950) encouraged one group of participants (primary school children) to use actual water jars but who were also offered pen and paper to first work out answers: These participants persevered in using the more complicated rule. A second group of participants (university students) were not offered pen and paper to work out solutions: the degree of perseverance was reduced. See Vallée-Tourangeau et al. (2011) for a more detailed discussion of the data reported in Luchins and Luchins (1950).

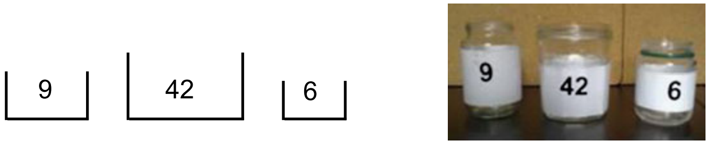


Fig. 13.2 The 4th training or set problem from the original Luchins procedure: The goal is to obtain exactly 21 units of liquid. The *left panel* shows how the problem was presented to participants on a sheet of paper underneath which they inscribed their answer. The *right panel* shows the set of three jars that were given to participants in the interactive group. Over a sink, they poured, transferred, emptied the water until they arrived at the desired amount. The solution is B − A − 2C or 42 − 9 − 2(6)

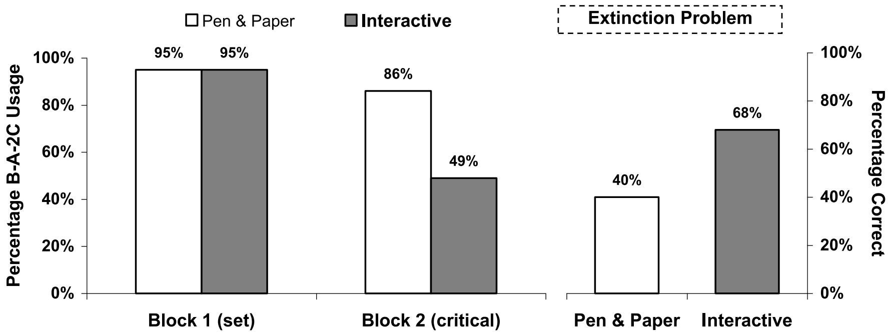


Fig. 13.3 Mean percent use of the B − A − 2C rule during the first five problems (the *set* problems) and during four out of the next five problems (the *critical* problems) in the Pen & Paper (*white bars*) and the Interactive (*dark bars*) groups—*left panel*; Mean percentage correct solution to the 8th problem (the *extinction* problem) which can only be solved with the rule A − C—*right panel*. Adapted from Vallée-Tourangeau et al. (2011, Experiment 1)

− A − 2C rule: The majority of their proposed solutions were the simpler rules involving either A + C or A − C. In turn, participants who worked on the traditional pen and paper version of the task still used, on average, the B − A − 2C rule on 86 % of the trials. This inability to employ or discover the simpler rule is well illustrated in their success rate at solving the extinction problem that could only be solved with the rule A − C (see the right panel of Fig. 13.3). A significantly greater proportion of participants solved this problem in the interactive group (68 %) than in the pen and paper group (40 %).

Mental set as examined with the original Luchins procedure is a well established phenomenon: Our participants completing the task with pen and paper persevered in using a more complex solution to a problem and a majority failed to discover the simpler rule to solve the extinction problem. The transition from the set problems to the critical problems was not signaled explicitly to the participants. The problems are familiar, perhaps routine at this stage, and hence the schema developed during the training set is triggered and applied to the critical problems. This schema guides the distribution of attention and problem solving resources; features of the new problems compatible with the schema reinforce its activation and hence increases its control over behaviour (Bilalić et al. 2008, 2010), naturally encouraging the use of the complex solution. In turn, participants for whom the task was embedded in a physical and manipulable environment were significantly less likely to employ the complex rule for the critical problems, and could more easily employ a simpler rule for these and the extinction problem. This suggests that features of the environment exerted some control of behaviour, competing with the schema acquired during the training set. The affordances offered by the actual jars attract attention and guide action, and participants in the interactive group were more likely to pick up the simpler solutions for the critical set. Problem solving, as a result, was more efficient.

Insight Problem Solving

Transformation problems such as volume measurement problems are structured in terms of a well-defined space of intermediate states linked by simple discrete moves, with the goal state clearly imaginable. Insight problems are different in that the goal state, or resolution, is initially not visible or imaginable. With insight problems most participants initially experience an impasse from which they may or may not emerge to formulate a solution to the problem. The impasse is experienced as a result of a problem representation that is driven by 'organizing assumptions' (Segal 2004, p. 142) that mislead the reasoner and prevent him or her from anticipating the solution. Overcoming an impasse is understood to be driven by a representational change that re-casts the relationship among the elements of the representation or that redefines the role of these elements. This representational perspective on insight has roots in Gestalt psychology (e.g., Wertheimer 1959) and has been formulated in information processing terms by Ohlsson in a series of papers (1984, 1992).

The initial representation of the problem is based on the manner with which the reasoner configures perceptual elements that compose the problem (how these elements are 'chunked') and reflects the reasoner's comprehension that recruits long term memory knowledge and expertise. Thus this initial representation structured by perceptual chunks and conceptual assumptions guides how the reasoner will attempt to solve the problem. However that guidance may also constrain and impede successful problem resolution. Certain assumptions of the problem representation may need to be relaxed in order for the reasoner to solve the problem. In addition, the segmentation of visual information into chunks is an important determinant of the ensuing problem representation and the ease with which a reasoner can solve the problem. Configuring that information in different perceptual chunks may also be an important mechanism of representational restructuring in solving insight problems. These two mechanisms, chunk decomposition and constraint relaxation, were explored in a series of elegant experiments with matchstick algebra problems by Knoblich et al. (1999). A matchstick algebra problem is a false statement written with Roman numerals. The problem is solved by moving (but not discarding) a sin-

Table 13.1 The four matchstick algebra problem types developed by Knoblich et al. (1999). Solutions for problems for Type A through C require relaxing constraints of increasing scopes, while solving problems of Type D involve decomposing a tight perceptual chunk

Type	Equation	Solution
A	VI = VII + I	VII = VI + I
B	I = II + II	I = III − II
C	III = III + III	III = III = III
D	XI = III + III	VI = III + III

gle stick such as to transform the false arithmetic statement into a true one. For example

$$\mathrm{VI} = \mathrm{VII} + \mathrm{I}$$

is a false statement that can be transformed into a true one by moving a single stick from the ‘7’ on the right of the equal sign to the ‘6’ on the left of the equal sign such as to yield

$$\mathrm{VII} = \mathrm{VI} + \mathrm{I}$$

Using matchstick algebra Knoblich et al. explored the importance of constraint relaxation and chunk decomposition in achieving insight. To test the importance of constraint relaxation, they developed three types of false statements the solution for which required relaxing constraints of different scopes (see Table 13.1). Solving Type A problems involved relaxing a relatively narrow constraint that numerals cannot be decomposed. Relaxing that constraint enables participants to transform a numeral to make the statement true. Solving Type B problems involved relaxing a constraint with a broader scope, that is one including the constraint on manipulating operators. Thus solving Type B problems involved changing the operator. Solving Type C problems involved relaxing a constraint with an even broader scope, namely the constraint that people rarely communicate in tautological terms. Hence to solve these problems, participants must realize that tautologies are acceptable. Knoblich et al. predicted that the solution rates of these three types of matchstick algebra problems would be a function of the scope of the constraint to be relaxed, with the narrow constraint of Type A problems the easiest to relax and hence to solve, and the broad constraint of Type C problems the hardest to relax and solve. Knoblich et al. observed the highest rates of problem solving success for Type A problems, followed by Type B problems, and the hardest problems were Type C.

These authors also tested the importance of chunk decomposition in solving these problems. They developed a fourth type of problems, Type D, by taking Type A problems with their narrow value constraint but used Roman numerals that are more perceptually complex, forming tighter perceptual chunks (see Table 13.1; in this example the solution involves decomposing the ‘X’ perceptual chunk into a ‘V’). Knoblich et al. predicted that problems of Type D would be harder to solve than problems of Type A, and that was exactly what they observed.

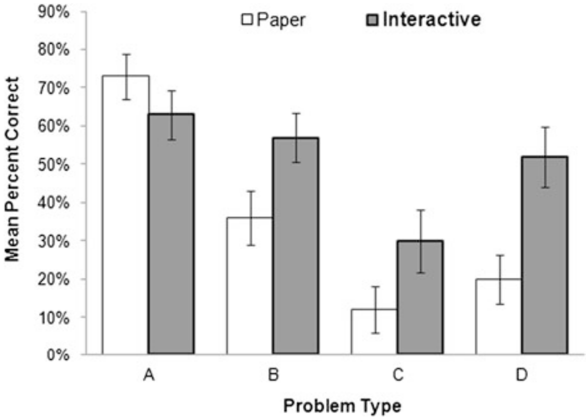


Fig. 13.4 Mean percentage of correctly solved problems of Types A, B, C and D in the paper group (*open bars*) and the interactive group (*dark bars*). *Error bars* are standard errors of the mean. Taken from Weller et al. (2011)

Interactive Matchstick Algebra As with the traditional water jar task, the matchstick algebra task is not an interactive one: Although the problems refer to concrete, common artefacts, a key feature of the Knoblich et al.'s experimental procedure is that participants are never invited to manipulate matchsticks as such in solving these problems. The so-called matchstick algebra problems don't involve actual matchsticks. Rather the false arithmetic statements are presented on a computer screen and participants voice their proposed solutions, which are then noted by the experimenter. Given the obvious in-principle manipulability of matchsticks in the 'real' world we investigated constraint relaxation and chunk decomposition using an experimental procedure where participants could physically touch and manipulate matchsticks in solving the problems (Weller et al. 2011). We designed[3] a magnetic board (27 cm × 21 cm) on which participants created and modified Roman numerals and algebraic statements using magnetized matchsticks (0.5 cm × 4.5 cm). As in Knoblich et al.'s Experiment 1, our participants were given 4 examples of Types A and B problems and 2 examples of Types C and D problems for a total of 12 problems. Participants were randomly allocated to one of two groups. In the paper group, an answer booklet was prepared where each problem was presented on a separate sheet of paper and for each problem participants announced their solution to the experimenter. In the interactive group, participants were shown the same booklet but for each problem they were asked to re-create the false equation on the magnetic board and then were asked to solve the problem.

The mean percentage success for each of the four types of problems in both groups of participants are plotted in Fig. 13.4. Knoblich et al. predicted and observed the following pattern: (i) A problems easier than B problems; (ii) B problems easier than C problems; (iii) A problems easier than D problems. (They did not formulate predictions concerning the relative difficulty of Type D problems vs. Types B and C problems.) Note the patterns of problem solving success in the paper group

[3]We thank Susan Cook for her help with the design and construction of the artefacts used in this experiment.

closely replicated the one predicted and observed by Knoblich et al. However, in the interactive group, solution rates for Type A problems were identical to the solution rates for Type B and D problems; C problems remained the hardest to solve.

Predictors of Success We also sought to identify the cognitive individual differences that predicted performance in both versions of the task. These data then can identify the skills and capacities implicated when reasoning in the head and when reasoning in the world. Thus, we profiled our participants in terms of IQ using the National Adult Reading Test (NART), which is a proxy measure for the WAIS IQ and in terms of visuo-spatial reasoning abilities using the Beta III IQ test. In addition we devised a short test of numeracy that all participants took before the start of the experiment. In the paper group, rates of successful solutions across all four problem types were significantly correlated with performance on the numeracy test. Performance on the NART was a significant predictor of performance in the paper group but not in the interactive group. Visuo-spatial abilities as measured by the Beta III IQ test significantly predicted performance in the interactive group, but not in the paper group.

This pattern of predictive relationships offers an important window onto distributed cognition. For the participants who could not interact with the roman numerals, their performance on the matchstick algebra problems was predicted by their numeracy skills and a measure of general intelligence, but not by a measure of visuo-spatial ability. The reduced interactivity meant that participants in the paper group had to rely on their internal/mental computational abilities to simulate certain matchstick movements. These mental simulations were always confronted by unchanging perceptual feedback of the false statement; their mental projections could not be externalized or reified to anchor subsequent mental projections and reifications. General intellectual skills could predict performance in the paper group because they capture in part people's mental abilities including executive function skills necessary to mentally simulate matchstick movements.

In turn, performance in the interactive group could proceed on a very concrete project-create-project (Kirsh 2009) interactive cycle: changes in the physical representation reified a hunch, which anchored subsequent projections which could in turn be translated in actual changes in the physical environment. It is for these reasons that we believe performance was significantly superior in the interactive group. Note, though, that performance on the Beta III visuo-spatial tests was significantly correlated with the matchstick algebra solution rates in the interactive group, but not in the paper group. This suggests that while participants could exploit the physical environment to help them think, they nonetheless relied on visuo-spatial skills to get the most out of interactivity: The higher they scored on the Beta III, the greater the number of matchstick algebra problems they were able to solve.

Problem Solving Research at a Crossroads

The data presented here with volume measurement problems and matchstick algebra were generated on the basis of interactive versions of these tasks that coupled thinking with artefacts which created dynamic problem presentations. And while these tasks remain constrained and artificial, we would argue that the interactive methodology employed offers a much closer approximation of real-world problem solving behaviour than the non-interactive static problem presentations originally employed. Clearly, the Gestalt notions of einstellung and reproductive thinking were important drivers of that original research, which also find purchase in more recent efforts (e.g., Bilalić et al. 2008, 2010). Similarly, Ohlsson's (1992) perspective on the psychology of insight encouraged a clearer elaboration of the mechanisms of representational change that led to testable predictions.

The data reviewed in this chapter validate the importance and heuristic value of conducting problem solving research from a systemic cognition perspective. Gestalt notions and theories of representational change that can accommodate thinking performance in decontextualised and disembodied experiments, must be extended and put to the test in interactive problem solving environments. Dynamic problem presentations offer a fluid set of thinking affordances not only to the participants in experiments such as those summarized here, but also for researchers who aim to develop models of thinking. Distributed problem representations reflect the recruitment and coupling of resources that are internal and external to the thinking agent. As a result, the control over behaviour is also distributed among internal and external factors. The significant reduction in mental set with the interactive water jar task indicates that the salient features of the perceived and manipulable objects overcome (indeed overwhelm) the familiar schema acquired during the training phase. This is strong evidence that a different process model of the water jar task is implicated in the interactive version. Similarly, the patterns in the correlations between test of cognitive abilities and performance in the matchstick algebra problems also converge on the notion that designing interactive versions of these tasks is not simply an exercise in making things more concrete to facilitate reasoning. Rather, the concreteness of these tasks and the necessary interactivity engage a different set of cognitive, perceptual and motor skills.

The psychology of problem solving is at a crossroads. It can continue to elaborate process models for two-dimensional non-interactive tasks, bolstered by neuroscientific evidence implicating specific brain regions (e.g., Geake and Hansen 2010). Or it can change direction, and seek to design experimental environments that are more representative of problem solving as situated, embedded, and embodied activities, that correspond to the manner people think and behave. To be sure, interactivity introduces a large number of degrees of freedom which reduce the psychologist's control over the experimental environment, but it also offers a much richer set of data from which to infer the reasoning mechanisms at play when solving problems, mechanisms that can better inform how reasoning outside the laboratory environment proceeds. In addition, interactivity and the distributed nature of thinking processes question the importance accorded to neuroimaging evidence both from a theoretical and practical point of view. The impressive spatial resolution obtained from some neuroimaging technology, such as with functional magnetic resonance imaging, is predicated on an immobile thinking agent who must *not* interact with his or her environment for fear of contaminating the imaging data.

The forces that will shape the future direction of problem solving research will likely reflect factors that have more to do with the sociology of science than with theoretical advances. Data gathered in the past 70 years of research on problem solving were mainly observed and recorded in thinking contexts that afford little or no interactivity. These methodological decisions are rarely questioned and represent orthodox scientific practice. The data obtained are taken to provide a representative window onto important reasoning phenomena—these become the canonical data that must be explained, for which process models are elaborated, and which inform the textbooks that educate new generations of psychologists. The evidence reviewed in this chapter questions these data and these methodological decisions. It invites psychologists to rethink how to study problem solving behaviour in laboratory environments and the nature of the process models that explain that behaviour.

References

Bilalić, M., McLeod, P., & Gobet, F. (2008). Why good thoughts block better ones: the mechanism of the pernicious Einstellung (set) effect. *Cognition*, *108*, 652–661.

Bilalić, M., McLeod, P., & Gobet, F. (2010). The mechanism of the Einstellung (set) effect: a pervasive source of cognitive bias. *Current Directions in Psychological Science*, *19*, 111–115.

Geake, J. G., & Hansen, P. C. (2010). Functional neural correlates of fluid and crystallized analogizing. *NeuroImage*, *49*, 3489–3497.

Giere, R. N., & Moffatt, B. (2003). Distributed cognition: where the cognitive and the social merge. *Social Studies of Science*, *33*, 301–310.

Gholson, B., Eymard, L. A., Morgan, D., & Kamhi, A. G. (1987). Problem solving, recall, and isomorphic transfer among third-grade and sixth-grade children. *Journal of Experimental Child Psychology*, *43*, 227–243.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.

Jeffries, R., Polson, P. G., Razran, L., & Atwood, M. E. (1977). A process model for Missionaries-Cannibals and other river-crossing problems. *Cognitive Psychology*, *9*, 412–440.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, *73*, 31–68.

Kirsh, D. (1996). Adapting the environment instead of oneself. *Adaptive Behavior*, *4*, 415–452.

Kirsh, D. (2006). Distributed cognition: a methodological note. *Pragmatics & Cognition*, *14*, 249–262.

Kirsh, D. (2009). Problem solving in situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 264–306). Cambridge: Cambridge University Press.

Kirsh, D. (2010). Thinking with external representations. *AI & Society*, *25*, 441–454.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*, 1534–1555.

Latour, B. (1986). Visualisation and cognition: Thinking with eyes and hands. In H. Kuklick (Ed.), *Knowledge and society studies in the sociology of culture past and present* (Vol. 6, pp. 1–40). London: Jai Press.

Lave, J. (1988). *Cognition in practice*. Boston: Cambridge University Press.

Luchins, A. S. (1942). Mechanization in problem solving: the effect of einstellung. *Psychological Monograph*, *54*(248), i–95. doi:10.1037/h0093502.

Luchins, A. S., & Luchins, E. H. (1950). New experimental attempts at preventing mechanization in problem solving. *The Journal of General Psychology*, *42*, 279–297.

McKelvie, S. (1990). Einstellung: Luchins' effect lives on. *Journal of Social Behavior and Personality*, *5*, 105–121.

Ohlsson, S. (1984). Restructuring revisited II: an information processing theory of restructuring and insight. *Scandinavian Journal of Psychology*, *25*, 117–129.

Ohlsson, S. (1992). Information processing explanation of insight and related phenomena. In M. T. Keane & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (Vol. 1, pp. 1–44). London: Harvester Wheatsheaf.

Pearce, J. (2008). *Animal learning and cognition* (3rd ed.). Hove: Psychology Press.

Segal, E. (2004). Incubation in insight problem solving. *Creativity Research Journal*, *16*, 141–148.

Simon, H. A., & Newell, A. (1962). Computer simulation of human thinking and problem solving. *Monographs of the Society for Research in Child Development*, *27*, 137–150.

Vallée-Tourangeau, F. (2013). Interactivity, efficiency, and individual differences in mental arithmetic. *Experimental Psychology*. doi:10.1027/1618-3169/a000200.

Vallée-Tourangeau, F., Euden, G., & Hearn, V. (2011). Einstellung defused: Interactivity and mental set. *Quarterly Journal of Experimental Psychology*, *64*, 1889–1895.

Vallée-Tourangeau, F., Guthrie, L. G., & Villejoubert, G. (2013). *Moves in the world are faster than moves in the head: interactivity in the river crossing problem*. Manuscript under review.

Watson, J. (1968). *The double helix*. Baltimore: Penguin.

Weller, A., Villejoubert, G., & Vallée-Tourangeau, F. (2011). Interactive insight problem solving. *Thinking & Reasoning*, *17*, 429–439.

Wertheimer, M. (1959). *Productive thinking* (enlarged ed.). New York: Harper & Brothers Publishers.

Chapter 14
Systemic Cognition: Human Artifice in Life and Language

Stephen J. Cowley and Frédéric Vallée-Tourangeau

Abstract Rather than rely on functionalist or enactivist principles, *Cognition Beyond the Brain* traces thinking to human artifice. In pursuing this approach, we gradually developed what can be deemed a third position in cognitive science. This is because, like talking, doing things with artefacts draws on both biological and cultural principles. On this *systemic* view, skills embody beliefs, roles and social practices. Since people rely on interactivity or sense-saturated coordination, action also re-enacts cultural history. Bidirectional dynamics connect embodiment to non-local regularities. Thinking thus emerges in a temporal trajectory of action that takes place within a space populated by people and objects. Utterances, thoughts and deeds all draw on physical, biological and cultural constraints. Even plans are shaped as first-order activity is shaped by second-order structures. Intentions and learning arise as dynamics in one time-scale are co-regulated by dynamics in other scales. For example, in ontogenesis, interactivity prompts a child to strategic use of second-order language. By linking cultural scales to inter-bodily dynamics, circumstances are coloured by resources that serve in using simulation to manage thought, feeling and action. The systemic nature of cognition connects *now*, the adjacent possible, implications for others and, potentially, social and environmental change.

All Together

The concept of thinking can evoke an isolated man who struggles with abstract problems on his own (curiously, the *penseur* is typically a man). The mark of such a

S.J. Cowley (✉)
School of Psychology, University of Hertfordshire, College Lane, Hatfield, Hertfordshire, AL10 9AB, UK
e-mail: s.j.cowley@herts.ac.uk

S.J. Cowley
Department of Language and Communication, University of Southern Denmark, Sdr. Stationsvej 28, 4200 Slagelse, Denmark

F. Vallée-Tourangeau
Department of Psychology, Kingston University, Kingston upon Thames, Surrey, KT1 2EE, UK
e-mail: f.vallee-tourangeau@kingston.ac.uk

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8_14, © Springer-Verlag London 2013

thinker appears in the furrowed brow of Rodin's famous statue. Yet empirical studies show that this is thinking of an unusual kind that is associated with an unusual type of task. While ideas can emerge in silence under a furrowed brow, thinking more typically arises as people battle with an invoice, look for the foreman or choose to trust a web site. Though people can think alone, they also do so when looking at X-rays, drawing geometrical shapes or, indeed, talking with others. In all cases, brain-side activity is inseparable from world-side events. Drawing on human artifice, thinking is co-constituted by speech, movement and gesture. People distribute control as they link routines, make instant judgements and coordinate as they act. Looking at and beyond mid-twentieth century functionalist models, the fine- and coarse-grained studies of *Beyond the Brain* show the sheer diversity of human thinking.

The metaphor of extended mind brings home that people use artefacts as they think alone and together. Extended systems can contribute much to thinking because outcomes can be separated from processes. Thinking can thus be viewed as separate from thoughts or, indeed, identified with reading *Wikipedia*. Wary of dangers arising from profligacy, we prefer to adopt Järvilehto's (2009) methodological goal of establishing how organism-environment systems achieve results. On this *systemic* view, anything that is perceived as an outcome can have cognitive consequences. Thus a wink or a digit (e.g., 42) may evoke, for example, the meaning of life.[1] Perceived products can change a cognitive frame or trigger shifts in roles, routines and the use of equipment. Shared or non-local resources often bring action—and thinking—under a degree of collective control. Bodily instruments use artefacts and artifice as people act and think. At times, actions are triggered by a situation; at others, they are authored as people behave with intent. In considering the triggered and the deliberate, emphasis falls on how human interactivity links cultural and other non-local resources. On the systemic view, therefore, much depends on coordination. From this position we conclude by considering both strengths and weaknesses of functionalist and enactivist cognitive science. Though thinking uses real-time processes, much depends on primates who draw on multi-scalar cultural resources.

Investigating (Human) Organism-Environment Systems

In Simon's metaphor, computational theories can be regarded as describing journeys in a problem space. In Chap. 3 of *Cognition Beyond the Brain* input-output models are applied to judgement aggregation (Spiekermann 2013) and, in Chap. 4, to computer-induced trust (Ben-Naim et al. 2013). However, such models overlook rapid or pico time-scales. In tens of milliseconds, dynamics use biology as parameters (e.g., syntax, social norms) impact on *now*. Thinking arises as people coordinate under physical, biological and cultural constraints. Artefacts can be representations

[1] In Douglas Adams's (1979) 'The hitchhiker's guide to the galaxy' a super-computer calculated the answer to 'What is the meaning of the life, the universe and everything?' to be '42'. In making this link, we treat a digital product as separable from some kind of process.

(Kirsh 2013, Chap. 10), prompts to solution probing (Steffensen 2013, Chap. 11), shapers of narratives (Baber 2013, Chap. 8), or used to prevent a (fictional) nuclear disaster (Jones 2013, Chap. 7). Indeed, as shown in the case study of a brain damaged person, persons normally draw on objects and events to create a stable mooring or a familiar world (Hemmingsen 2013, Chap. 6). Further, the dynamics are bidirectional. Even brain-side habits and anticipatory control draw on working afferent and efferent nerves. So how is thinking managed across the skin and sensory organs?

As cognitive science came to acknowledge that cognition is embodied and embedded, the functionalist paradigm was challenged by enactivism (see, Thompson 2007a, 2007b; Stewart et al. 2010). Rather than take sides in this debate, we turn from models of input-output relations to "start from the determination of the results of behavior" (Järvilehto 2009, p. 118). By so doing, we seek out the system's constituents that determine "the achievement of these results" (Järvilehto 2009, p. 118).[2] Our *systemic* perspective asks *how* thinking self-organises across individuals, dyads and groups. Configurations of control change as brains, persons and artefacts are put to use under shifting supra-personal constraints. Results arise as parts trigger effects in larger systems and, of course, systemic wholes constrain the operations (and modes of organization) of these parts. Without having to depend on phenomenology, the approach recognises that human observing is the key to understanding extended systems. If they so choose, people can find (and create) links between stars and tornadoes or molecules, frogs and group decisions. This happens because, in a cultural and physical world, skills are honed by using culture to develop the habits that make us singular individuals.

In the domain of the nonliving, events obey principles of dynamics and/or design. This is in stark contrast to the world of biology. Emphasising this *epistemic cut*, Pattee (1996) argues that organisms set parameters to measure and exert (a degree of) control over dynamic change. Where successes endure, the effects give rise to an ecosystem. Evolution thus gives rise to lineages of more or and less connected organisms. Gene-culture coevolution (e.g. Lumsden and Wilson 1981; Gintis and Gintis 2011) further augments the complexity of human modes of life. In taking a systemic view of cognition, the focus falls on agents whose actions and sense of self are partly constituted by engagement with institutions, artefacts and practices. Each and every singular person emerges from a history of events that binds the cultural with the biological. Like plants, slime moulds and ants, living humans are both parts of supra-personal systems and unique self-assembled wholes. In what follows, however, our emphasis falls on species-specific mechanisms (and their parts) that sustain the heterogeneous phenomena that contribute to human thinking.

Unlike other species, humans aggregate judgements, construct roads and overcome impasse. While individual actions matter, much depends on cooperation and competition. In spite of this, philosophical tradition has usually construed thinking as an 'inner process'. Ultimately, the view probably derives from human affinity for

[2]We diverge from Järvilehto slightly in that his focus is on constituents of the living system. We, by contrast, are equally concerned with the role played by historical and nonliving parts.

locomoting systems. Since self-motility co-evolved with neurons and central nervous systems, it is tempting to assume that thinking is wholly controlled by brain-side events. Indeed, neurons differ from other cells (and cell-assemblies) precisely in that they alter other neurons' activity and metabolic conditions: there is some central control. While bacteria use biophysics to move and grow, embrained creatures like ants and fish reliably discriminate action trajectories from results (Järvilehto 2000). A selection history and, often, learning attunes functional information with action.[3] Thus, an ant may select pieces of earth for a nest or a fish synchronise its movements to escape from predators. With brains, perception and learning give rise to mechanisms that integrate structures with different evolutionary histories. In vertebrates (and many invertebrates) organisms manage both self-directed behaviour and anticipatory action. Nonetheless, many results are partly dependent on world-side objects and events.

Anticipatory human behaviour is illustrated by experimental work on gaze in reading aloud (Järvilehto et al. 2009). Saccading is often far too rapid to use input-driven processing. At times, it follows the onset of voicing; we look to confirm expectations. Far from decoding inscriptions, readers exploit what they see by anticipating and monitoring their gaze. This remarkable fact ensures that human thinking can be constrained by historically derived constructs. In developmental time, these can alter what is expected. Thus, users of computers rely on non-local constraints to develop anticipatory habits. The information processing of computer games encourages interactivity. While further discussed below, human interactivity can be defined as *sense-saturated coordination that contributes to systemic function*. Gamers engage with an extended system to connect the affective, the normative and the habitual: anticipatory dynamics emerge. As embrained organisms, they rely on sense-making; however, as shown in Chap. 5, cellular systems have counterparts (Markoš et al. 2013). The chromatin mechanism is part of a natural technology that anticipates supra-cellular needs. To understand human interactivity, however, much is gained from attention to well-engineered systems. As Giere (2004) suggests, the Hubble telescope offers a paradigm example of how human understanding relies on the world beyond the brain.

By orbiting in space, the telescope significantly extends the cognitive powers of skilled observers. Its network of causal relays (hardware and software) is designed to make otherwise invisible celestial light available to earth-bound humans. The telescope generates digital output which is reformatted for the eye. A human observer is moved to reiterate saccading by images that call up singular knowledge/experience.

[3]Following Sharov (2010), functional information can be traced to "a set of signs that encode the functions of the organism" and, in addition, of "signs that control the functions" (1058). Functional information thus includes genome, epigenome, internal messengers (e.g., mRNA, miRNA, transcription factors, kinases, and phosphatases), external messengers (e.g., pheromones), and natural signs (e.g., temperature and salinity of water). Moreover, the notion of functional information also applies to artificial signs (and what agents treat as signs). This view is narrower than Pattee's characterisation of living systems in relation to the self-organization of parameters or symbols that measure and control their dynamics. In evolution, symbols often lose their functions; however, signs are events that an agent uses/interprets in functional ways.

As a result of how looking organizes neural activity (and vice versa), inferences can be made about, for instance, a distant galaxy. Arguably, sense-making in human infants follows a similar logic. Together with caregivers, they enact extended cognitive systems (see Cowley 2003, 2004). Like stargazers, infants learn to separate self-generated actions/interpretations from outcomes. Just as with Hubble images, these serve to bring about informative results. As a baby gains skills in attending, discriminating and talking, its brain shapes experience of a familiar world (Kiverstein and Farina 2011). Living with extended systems permits cultural constraints to influence experience and understanding. As children discover real and imagined situations, they also learn to anticipate what others feel, think and do. The same logic applies in evolutionary time.

For Donald (1991, 2001) human brains were transformed by cultural activity. People became media sensitive as, using mimetic skills and social organization (Cowley 2012), artifacts and routines came to be managed around experience. Abilities for anticipating arise in linking circumstances to a grasp of events. Long before language, humans developed cultural-cognitive networks (Donald 2007). Whereas all primates experience recurrent situations, human ontogeny unfolds as people co-construct *events*. Human life stories link social perception, action and, later, verbal thinking. Crucially, individuals live a now that integrates structures with many origins: immediate experience uses cultural time-scales. Perception meshes with action as people monitor movements, effects and how these fit expectations. Perception drives action; action drives perception: embrained systems use what Berthoz (2010) calls *perçaction*. Habits and procedures allow humans to develop strategic action. Brains and cells, moreover, use similar motifs or principles such as using statistical information and reciprocal excitation/inhibition. For Berthoz (2012), these exemplify *simplexity* which, not surprisingly, reappears in human action. For example, varying and reiterated vocal patterns (Blair and Cowley 2003) link circumstances with talk. As in most intelligent activity, a great deal depends on inhibition. Simplex principles are selected because systems use competition, decision-making and stability (making similar choices in many circumstances).

Bodies Together: Living and Nonliving

The metaphor of extended mind licences a return to viewing the public world as (partly) *constitutive* of language and thinking. For Peirce (1940), Mead (1932), Dewey (1896) and others, thinking consists in cognitive or semiotic relations. Similar ideas appear in phenomenology and, of course, Wittgenstein (1958) and Heidegger (1971). In philosophy, brains (and minds) attract more attention than dynamics—and especially inter-bodily events.[4] Like Chomsky (1957) and Dennett (1991), extended functionalists such as Clark (1996, 2008) and Wheeler (2004) treat

[4] Stuart (2010) invokes enkinaesthesia or felt bidirectional coupling between organism and environment. Since it applies to most (perhaps all) living species, it leaves out non-local phenomena such as language and culture.

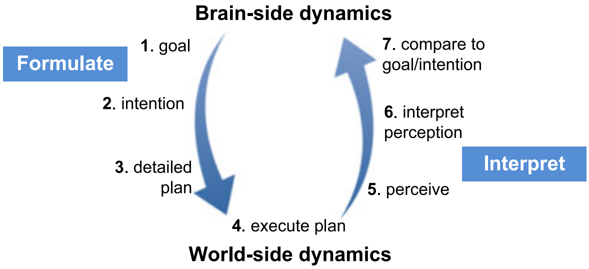


Fig. 14.1 The interactivity cycle (as in HCI)

language as verbal. By identifying it with linguistic products ('material symbols'), they reify second-order constructs. From a systemic perspective, they use naïve realism to identify human perception with how von Neumann machines process (Shannon) information. In fact, even language is based in sense-saturated activity. In Thibault's words, "A growing body of evidence shows that interactivity, not abstract symbol manipulation, content transmission, or information processing centred on the internal mental processes of the individual, is the key to human learning, cognition and intelligence." (2011, p. 13). Language connects a non-local domain of cultural patterns (e.g. beliefs, words and numbers) with physical events in time. It is thus part of perçaction and, as such, language shapes a sense of now—of what can be perceived and done. Like other forms of cognition, it is grounded in interactivity which, as shown by several papers in this volume, also serves in solving problems *without* speech. In dealing with objects, actions are constrained by familiar cultural constructs. Interbodily results thus affect both the observed and acts of observing. As Hacking (1999) shows, many cultural products arise from a history of social action. In an extended system, sense-making depends on the workings of what Järvilehto (2000) calls *joining organs* (e.g. eyes, brain, hands, skin). As with the Hubble telescope, these co-function as a person links cultural knowledge to causal relays. Dynamics allow products of earlier behaviour to transform later events (often managed by other people). Interactivity can thus be used to elicit expected outcomes. Anticipation has the same basis in a computer game. Designers exploit a propensity for perseverance and practice to ensure that players' habits include pressing 'restart'. It is not surprising, therefore, that HCI was the first field to recognise the importance of sense-saturated coordination (e.g. Kiousis 2002). This is, above all, because interactivity cycles shape human anticipation. Using an amended version of Kirsh's (1997) model, this can be visualised as in Fig. 14.1.[5]

Computers encourage users to develop routines. Given insistence on sequences, actions are broken into moves that are called *formulating* and *interpreting*. Visible screen changes shape perception (as input) and a user's actions (or output). The visible can lead to a material interpretation. For a user, the extended system's output (which results from a person's action) has meaning-potential. This functional

[5]While this artificial kind of interactivity pertains to an individual's thinking—not processes between people—it shows how a machine can sponsor human actions.

outcome invites monitoring because it not only offers feedback but, if noted, unexpected outcomes can be reversed. Over time, interactivity cycles come to be completed without intensive monitoring. Habits build cultural products into interpretation as the user gains a sense of the simulated situation. This serves in deciding what to do next (in various time-scales). For this reason, the interactivity cycle serves as the grounding of all human action.

In connecting cultural products with lived experience, people depend on perceiving outcomes against a trajectory of events. Where reinforced by rewards and/or user commitment, skills become situation-specific. However, in human-human encounters, we find neither HCI-style sequentiality and directionality nor its reversibility. In caregiver-infant encounters, for example, interpretation *is* formulation (and vice versa). Far from being reversible, human interactivity becomes an unending two-way flow. Nonetheless the HCI model brings home a simple truth. Because people distinguish between outcomes and the results of their behaviour, interactivity engenders both statistical and more self-conscious learning: outcomes act as markers for the pertinence of cues. In different circumstances, a doctor's concerns can be aroused by patterns on an X-ray or a robust sound can inject sense into wordings. Perceptual capacities link outcomes, habitual modes of attending, and an individual's anticipatory skills.

Interactivity binds procedures from the slow time-scales of culture with inter-bodily dynamics and, thus, a person's sense of circumstances. Second-order constructs thus inform situation relevant skills, interpersonal procedures and, with time, an individual's knowledge configuration. Bidirectional dynamics train up bodily action to link one person's concerns with those of others. Interactivity may be used to taste porridge before adding salt or in solving geometric problems (Kirsh 2013). Inconclusive results lead to reiteration, variations on a theme or, perhaps, acting differently (e.g., adding sugar). The pattern appears in the laboratory task where problem solvers are asked to assemble a cheap necklace. In an experimental setting, as Fioratou and Cowley (2009) show, problem solving strategies vary between groups who represent the problem graphically or handle a real necklace. Not surprisingly, perhaps, interactivity raises both efficiency and solution rates (Fioratou and Cowley 2009). In Chap. 13, Vallée-Tourangeau and Villejoubert (2013) report comparable results on, for example, well-known matchstick arithmetic problems. Turning to the film *Crimson Tide*, in Chap. 7, Jones (2013) shows how interpretation and formulation connect bodies, instruments, cultural techniques, verbal duels and, interpersonal antipathy. Steffensen (2013) shows, in Chap. 11, how ritualised 'rejection' of a document prompts shared insight. Action emerges under social constraints as parties 'from a feeling to figuring out what to do.' They link sense-saturated events with non-local concerns or, prosaically, fears about non-payment of the invoice. Just as in a computer game, events induce functional reorganization. Material objects prompt individuals to invoke past acting, feeling and speaking. Like the slow processes of adaptation, development and learning, interactivity shapes skills, strategies and understanding. Though probably based in social encounters, its scope is extended by nonliving systems. With experience, motivated actions use cultural criteria in ways that become remarkably differentiated.

Distributed Control

Bidirectional dynamics connect sense-making with actions, words and feelings. As people shift between roles, human behaviour appears adaptive, dependent on learning, and, compared to other primates, flexible. Much depends on, not what a person knows, but skills in distributing control. Indeed, Froese and Gallagher (2012) argue that what is usually explained by Theory of Mind can be traced to a history of interactions. If so, it is surely because, as in infants (see, Spurrett and Cowley 2004) understanding arises during events that use human interactivity. It is therefore important to clarify how the supra-personal influences behaviour. In using the Hubble, human perceptions (e.g., seeing a galaxy) can be traced to attributes that use the system's outputs (e.g., a pixelated image):

> A distributed cognitive system is a system that produces cognitive outputs, just as an agricultural system yields agricultural products. The operation of a cognitive system is a cognitive process. There is no difficulty in thinking that the whole system, no matter how large, is involved in the process. But there is also no need to endow the whole system with other attributes of human cognitive agents (Giere 2004, p. 771).

Interactivity drives anticipatory action because perçaction is experienced as separable from systemic output. For Giere (2011), this 'fundamental asymmetry' allows the human agent to make the most of an extended system. While this uses human attributes, its causal relays rely on input and produce output that is separate from interactivity.[6] Indeed cognitive extensions gain much power from the 'separability' of the whole system from (current) human concerns. Moreover, this *principle of cognitive separability* has a parallel in the evolution of brains. Just as embrained systems distinguish behavioural results from a trajectory of action, extended systems allow material markers to accrue cultural values. By cooperating in making stone tools (mimesis), a person may come to 'see' how to knap a given flint. On this basis, perçaction attunes to historically derived functionality.

Vocal and bodily gesture may also exploit the process to establish 'replicable constraints' (Pattee and Rączaszek-Leonardi 2012). Whether or not this is the case, one cannot doubt that modern humans use habits and patterns as affordances. The principle of cognitive separability as what allows a wink or digital output (e.g., 42) to be used in reframing construal. It is because eye-closing can be seen *as* a wink or 42 used *as* a reminder that its materiality has meaning-potential. Much learning can thus result from using objects, following procedures, and adopting kinds of organization (e.g., styles, value systems). Many human powers depend on the evoked or, in Gibson's (1979) terms, the 'education of attention'. Given a cue's real-duration,

[6] "Nevertheless, there remains a fundamental asymmetry in my view. I want to call a system cognitive because it produces cognitive outputs, but refuse to call it knowledgeable because it produces knowledge. For me, the latter makes as little sense as calling a system edible because it was designed to produce edibles" (Giere 2011: 397).

just as circumstances disambiguate the perceived, interactivity can serve to reconstrue circumstances. We perceive objects/events as signs, see pictures, and create behavioural markers (Kirsh 2013). Like causal relays, these can change later construal. Cognitive separability shows its power in *loosely coupled systems*. As explained in Chap. 9, construction teams use 'all available resources' including the office layout (Perry 2013). In this sense, many decisions use supra-individual intelligence or, in Perry's terms, are *demand led*. However, though circumstances and culture always influence thinking and action, human initiative or authorship often come to the fore. In the wild, as in the laboratory (Evans 2010), different *kinds* of thinking occur.

Given the principle of cognitive separation, demand led perçaction can give way to more individually-centred thinking. Once again simplex principles apply. At a molecular level, distributed control uses the chromatin mechanism to regulate DNA expression (see, Markoš et al. 2013). Markers on histone tails act as *separate* structures that enact slower functional dynamics. In principle, this is like how a gamer develops habits and, if he recognises them, can act to change them. Complex effects arise as from competition across time-scales: while users may be persuaded to trust a given supra-personal system, it may also collapse. Banks, businesses and governments lose credibility when instabilities emerge in faster scales. In development, competing dynamics affect learning. As shown in Chap. 2, to become rational, children must resolve a paradox (Neumann and Cowley 2013). In achieving the necessary independence, reliance on a caregiver must be replaced by strategic following/violating social norms. Reward-getting skills enable children to cope with/exploit social situations. They come to rely on population-level regularities or, in short, a space of reasons. Given that this applies to populations, it can be well captured by formal models. Likewise, judgement aggregation gains from formalization—provided that crucial (functional) information is shared. Conversely, models of reason overlook how *individual* agents sponsor and author events: they leave out how people use organic time-scales to eliminate or inhibit potential information. In James's (1890) phrase, they cannot clarify *the thinking that goes on*.

Increased interactivity can index lack of functional information. In Chap. 13, Vallée-Tourangeau and Villejoubert (2013) show a correlation between use of external resources and lower 'on-board' competence. How a task is approached depends, in part, on individual competencies. However, while predicting an extended system's performance, this throws little light on interpretations. Apparently, while external parts influence outputs, human brains and interactivity frame situations, distribute control and permit/drive selection of cognitive strategies. If demand-led, these are often social or mimetic. Yet, as Steffensen (2013) shows, face-to-face encounters also prompt solution (or problem) finding. Extended systems enable people to recalibrate how they grasp an issue. A major cognitive role thus falls to repetition with variation that, for Wallot and Van Orden (2011), uses neural functions on the edge of chaos. Indeed, this may clarify why high uncertainty increases the use of affect/arousal. By hypothesis the role—and energy—of pico-scale activity diminishes in conflict-free settings. Using different control systems across time-scales may well derive from a person's familiarity with the cognitive ecology.

Neural resources enable people to use the environment to create a stable now. Given a sense of sponsorship (Hemmingsen 2013) human agents develop other modes of action. Anticipatory dynamics thus make the lived present part of a familiar world. In seeking balance, interactivity enables social practices to constrain human sense-making. As shown in the case study of CW (Chap. 6), external resources invite body-based control. Strikingly, this echoes what formalizations fail to capture—how we eliminate or inhibit possibilities. Further, sponsorship is extended by acting to author events. As Jones (2013) emphasises, on occasion, individuals seek to take control (and resist such attempts by others). In *Crimson Tide* both rebellion and resistance depend on subtle use of cultural procedures and products. People draw on explicit reasoning and, indeed, hypothetical thinking. Accordingly Jones, contrasts this with the compliance which is often emphasised in distributed cognition (e.g. Hutchins 1995). In shorter time-scales, it seems, anticipation serves to deal with the familiar. Yet, in the longer term, one can use a sense of disquiet in inhibition, rebelling and pursuing supra-personal change. People draw on skills with working memory that serve so-called type 2 thinking (Evans 2010). Experience and interactivity favour a literal approach to verbal constructs. By using collective resources, people author actions.

The Duality of the Living

Flexible and adaptive behaviour uses cognitive systems that draw on physical, biological and cultural constraints. Cowley and Vallée-Tourangeau's (2010) *sociocognitive* view thus encompasses both what people usually call 'thinking' and also cases like judgement aggregation or the use of systems that induce trust in cultural products. Such a view has its dangers. In Chap. 8, Baber (2013) stresses how the individual contributes to legal process and, in Chap. 7, Jones (2013) links moral responsibility to action. As with theories of extended mind, task-based models blur the details of what actually happens. To avoid fuzziness, one can ask how people act to control feeling, thinking and acting as they contribute to extended systems. Using the principle of cognitive separation, they treat cultural (and other) products as independent of later events. In learning to anticipate, people use structures originating in many time-scales (e.g., lips, affect and wordings). While partly demand led, thinking builds on individual skills and resources.

In examining human agency, another surprise emerges. Although functionalist approaches invoke intentionality, they best describe populations. By modelling trends rather than individual behaviour, they make no attempt to discriminate between ways of accomplishing tasks. Indeed, this is why Giere's fundamental asymmetry is so important. It ensures that extended processes generate outcomes that carry rich reserves of meaning potential. Even simple cultural products (e.g., 42) influence later events as the 'same' functional information is put to various ends. Far from relying entirely on physical invariants or statistics, living beings use valued experience: sense-making exploits a lineage's evolutionary history. The systemic perspective thus uncovers a strange duality. From e. coli bacteria to human star-gazers, living systems use functional information that is bound up with bidirectional dynamics. In a human life-world, much importance falls to constructs with a cultural history.

In Humberto Maturana's (1975) terms, observers exploit a consensual domain. They recognise and re-evoke coordinated outcomes to connect first-person experience with reality. Bringing this view to the distributed language movement (see, Cowley 2007, 2011), Kravchenko (2007, 2011) emphasises that a person constructs her or his own understanding. An individual's whole-bodied or first-order activity gradually becomes entrained with strategic ways of using second-order constructs (see, Love 2004; Thibault 2011). Human meaning-making arises from creating and construing signs that, in history, relied on combining phonetic activity with visible expression (in languaging). While recognising this, Maturana downplays bidirectionality. Rather than turn to how the social and material worlds influence living agents, he emphasises structural coupling or how the organism adapts what it does to ever changing perception. Culture and selection are thus marginalised. Although appeal to *languaging* re-establishes continuity across species, it fails to distinguish between sponsored and authored actions (Hemmingsen 2013). This is because radical constructivism cannot distinguish between, on the one hand, meaningful (first-order) activity and, on the other, more deliberate use of second-order constructs. By contrast a focus on interactivity allows the flow of action to produce (what is perceived as) output that affords opportunities for construal/action. When using cultural resources, a distancing arises when rapid 'old' brain processes are complemented by slow ones based on principles like inhibition (and skills in re-membering or re-acting).

Similar simplex principles appear in molecular systems. Thus DNA both constrains variation and permits the exuberance of metabolism. In Sharov's (2010) biosemiotics, functional information makes a total system viable. As Monod (1966) saw, e. coli bacteria can enhance their viability through a striking ability to discriminate between, say, suchrose concentrations. Since how and when this is done cannot be explained by physical laws, Pattee (1996) ascribes it to the epistemic cut between how living and nonliving systems control their dynamics. Whereas physical change draws on pre-extant boundary conditions, living systems set their own constraints (using measures or symbols). To self-regulate biosystems separate dynamics from controlling parameters. In a given life-world, inner and outer regularities take on functional value. The system's metabolism thus realizes values (for itself). Since boundary conditions evolve in different time-scales, inter-scalar processes demand bidirectional (or polydirectional) regulation. Thus, the same simplex principles appear in biology, language and cognition. Not only do people use the results of past behaviour to change later action (by strategy, habit and design) but they also create procedures, institutions and artifacts. These permit anticipation based on linking material regularities with non-local values. A person's world thus extends biology. The human niche uses both the various phenomena that shape thinking and what people themselves believe thinking to be or, in Hacking's (1999) terms, the locally dominant 'idea of thinking'. Our lives connect up situated events with non-local sense based in our modes of life.

Empirical Cognitive Science: A Third Position

People use resources that extend both across the skin and in historical time. Beyond the brain, they use bodily parts, interbodily resonances, artefacts, computational outputs and a history of cultural procedures and human ways of life. Not only is cognition embodied and situated but performance draws heavily on non-local patterns. By investigating biological function in terms of results, it becomes possible to specify how parts, procedures and modes of organization use simplex tricks like inhibition. Further, in stepping back from functionalist and enactivist views, it is possible to revisit old debates from a new perspective. In this modest sense, evidence based systemic work represents a third position in cognitive science.

While bodies and brains contribute to 'thinking', this also applies to physical, interbodily and cultural resources. Indeed, linguistics, psychology, anthropology and ergonomics (for example) can ignore many philosophical problems. Minds do not tackle problems; rather, people act within self-organizing aggregates that sustain supra-personal systems. Such events can be investigated independently of debates about what causes/constitutes cognition, the mind/body problem or mental states/representations.[7] While the folk speak *as if* minds used representations, living systems rely on functional information. Even if much of what people feel, think and do is demand led, individuals *also* author actions. Dynamics in one scale can be regulated by and regulate dynamics in other scales. Since the principle applies to language, life and cognition, the phenomena elude any single model—computational, dynamic, enactive or semiotic.

Cognitive Cycles

Viewed as process, mind can be reduced to strategies, habits and stimulus-based learning; alternatively, it can be expanded by input-based models of hidden process. Yet linear processing cannot describe even cellular dynamics. As argued by, among others, Gibson (1979), Bernstein (1967), Peirce (1940), von Uexküll (1926) and Dewey (1896), cognition is fundamentally cyclical. In rejecting the input-output models of functionalism, there is a close affinity between the systemic view and enactivism. However, far from treating human cognition as action-centred, it has been emphasised that activity is strictly multi-scalar. While motility is brain-based, the principle of cognitive separation makes much of what we do dependent on the world beyond the brain. Humans have been transformed by an evolutionary history of using extended systems. Second-order constructs allow people to distance actions from the promptings of the world. Language and cognition thus possess a duality

[7] For example the authors differ on the status of brain-side representations: while one of us sees no reason to challenge these, the other thinks that appeal to neural representations is mistaken. However, on our view of cognition, this matters little. Even brain-side, it may be a matter of the level of description.

that, once again, draws on simplex principles. While human thinking and language can be demand-led, interactivity is associated with constructs that draw on non-local constraints.

There is a fundamental asymmetry between an (extended) system that shapes human beliefs and knowledge and how individuals perceive/process functional information. In development, the skills draw on the structures of institutions, artifacts and languages. Although people end up conforming with their fellows, they can also modify resources. Indeed, perception of cognitive separation may favour digitalization (or information compression). Thus, formalization not only characterises macroeconomics but it also applies to grammars and judgement aggregation. Further, causal relays affect group and interpersonal activity. As reported in Chap. 12, novices gain from tracking seemingly arbitrary lines on an X-ray (Ball and Litchfield 2013) and, as shown in Chap. 8, software can be used to collect crime scene evidence in support of the prosecution (Baber 2013). Even if formal models leave aside bodily dynamics and how we select potentially valuable information, this is but one part of human cognition. Much depends on scales that reach beyond (and beneath) lived experience.

Marr's (1982) metaphor of cognitive *levels* offers a related view. While a task can be defined at a computational level, it draws on (lower) algorithmic and implementational levels. In these terms, functionalists and enactivists not only examine different tasks but, as has been noted, appeal to different kinds of granularity. Whereas functionalists focus on an algorithmic 'level', enactivists turn to bodily 'implementation'. This demonstrates lack of agreement about what cognitive science can/should investigate: while enactivists focus on first-person experience, functionalists examine what people (and brains) are *likely* to do. This contrast echoes nature's strange duality: whereas algorithmic models apply to populations (including neural populations), implementation uses bodily and inter-bodily relations. Perhaps this is because algorithmic models use folk descriptions of tasks (which are based in these terms) as opposed to observation of bodily or experiential activity.

Applied to language, functionalists adopt what Linell (2005) calls *written language bias*. On what Steffensen (2011) rightly dubs a conservative view, Clark (1996, 2008) views language as arrangements of (alphabet-like) 'material symbols'. Like Chomsky (1957) or Dennett (1991), he ignores interactivity and dialogue, underplays culture and ascribes linguistic cognition to the brain. While espousing a dynamical view, Wheeler (2004) ends up embracing a representationalist theory. One problem with functionalist models is that, in principle, they separate causal effects from our sense of *now* or what Hemmingsen (2013) calls first-personhood. By separating language from bodily dynamics, human skills are subordinated to the cognitive potential of Clark's 'ultimate artifact'. For similar reasons, no functionalist model can capture how interactivity contributes to problem-solving or brains generate a sense of familiarity. Much depends on the enactivist's lived time-scale; much, however, also depends on time-scales that pertain to populations of neurons or human agents. In such cases functionalist models carry considerable merit: the baby must not be flushed away with the bathwater.

Varela et al. (1991) focus on how minds come to terms with first-person experience. To explain the network of the brain, they turn to—not the social meshwork sustained by culture—but first-personhood. In linking the perspectives, the systemic view opens up both how the observer and observable phenomena arise (in different time scales). In pursuing a complementary view, it is clear that more is needed than a 'consensual domain' of material and social constructs. Indeed, making autopoiesis the basis for mind overplays the scope of human autonomy. Leaving aside the importance of historically derived attractors (or replicable constraints), language reduces to action (or techniques). Although it is entirely appropriate to emphasise the observer, enactivists overlook what lies beyond an organizationally closed organism-world. By beginning with the networks of the mind, not the ecology, second-order dynamics cease regulate or, indeed, be regulated by lived experience. Enactivists thus struggle with the meaning potential of artefacts, institutions and verbal patterns. While Bottineau's (2010, 2012) important work traces linguistic action to (culturally specific) linguistic techniques, he leaves aside how these are entrained by the lexicogrammatical patterns that Thibault (2011) has termed 'future attractors'.

The Hubble telescope shows the limits of both functionalist and enactivist models. While the observer's 'human attributes' (roughly, experience of seeing) are open to enactivist views, these throw little light on how people construe pixelated images (or utterance-activity). Such complexity does not reduce to interaction—or an individual history that produces a consensual domain. Ironically, the need to move beyond interaction shown by Auvray et al.'s (2009) demonstration that inter-bodily activity is regulated by patterns that are beyond an individual's awareness. If the insight that we can understand statistically is supplemented by recognition of the principle of cognitive separation—and ability to make sense of objects—this opens up the domain of language and the thinking of Evans's (2010) 'new brain'. Imbuing the observed with meaning-potential can be used to decide between competing demands, deal with conflicts and, of course, serves to discover/create new non-local resources. A focus on phenomenological and, in experimental settings, the statistical, underplays how pico-scale movements are integrated with experience and biocultural products. Too little attention has fallen on the supra-individual world or how gestures (and syllables) are produced. Without a systemic perspective, it is hard to link events across the skin with how extended systems prompt rapid shifts in first-personhood and interbodily control.

Regulating Inter-bodily Flow

Distributed cognition permits use of an asymmetry between acting in an extended system and taking a perspective. This is Giere's insight—we perceive events/objects in the world as separate from actions: we can detach from cognitive outcomes and even products of our own speech. Any noticeable contingency may be used to construe circumstances and/or, to reframe a cognitive context. In tightly controlled cognitive systems such as the cockpit of a plane (Hutchins 1995) or the game of Tetris (Kirsh and Maglio 1994), artefacts take on meanings as, for example, pilots stabilize decision-making. In looser systems, radical implications emerge. This is because material and/or phenomenal contingencies channel idiosyncratic interpretations. People use temporally and spatially extended regularities. Indeed, repeated vocalizations and tool-making movements can be modified and modulated in ways that result in social affordances. The most important insight of distributed cognition is, echoing Hollan, Hutchins and Kirsh, that cognitive processes can be distributed through time such that the 'products of earlier events can transform the nature of later events' (2000, p. 176).

In extended systems, orientations are constantly shifting. At one moment 42 is a number and, moments later, an allusion to the meaning of life (or a novel). To subdue disquiet, we develop habits and sensitise to non-local constraints. Perceiving constrains action as people manage activity and awareness. Social affordances connect the bodily with the normative. Anxiety and tension can be *about* feared non-payment of a debt; seeing muddy boots may *answer* a question. Interpretations build on a tightly constrained cognitive separability—we are not overwhelmed by freedom. While intrinsically meaningless and not the basis for behaviour, populations make consistent use of non-local resources. Patterns that are selected by dint of using a technique become constraints on future action. As a consequence, it is often enough to act in demand led ways. However, human thinking can also use the power of authorship.

Under (Partial) Collective Control

In a species where groups, dyads and individuals exploit self-organizing aggregates much depends on cognitive control. At times, this is more demand led; at others, it is looser and individual-focused. Control thus depends on management of the body, various second-order constructs, and lived experience. Not only is systemic output separable from actions but, just as strikingly, we offload information onto extended systems. Non-local patterns insinuate themselves into individual lives. In parallel, molecular systems use something like a *principle of metabolic separation*. To function under (partial) collective control, they evolved organic codes or procedures that draw on fixed parameters such as those used in protein synthesis (Barbieri 2009). However, as shown in Chap. 5, more complex processes associated with the chromatin mechanism are needed to manage the repair and regulation of metabolism: DNA expression thus serves higher-level systems (Markoš et al. 2013). Rather as DNA links physical structure with inert patterns, culture uses virtual structures that have a symbiotic relation with bodily and inter-bodily powers.

In Ross's (2007) terms, human uniqueness depends on the ecology. Just as microeconomics are under (partial) collective control, language works in parallel (Rączaszek-Leonardi and Cowley 2012). Contingencies serve to manage situations: cognitive separation imbues certain objects/events with meaning-potential. Ceasing to be part of the fabric of the world, patterns become signs (and wordings) that are replicable and, thus, available for later sense-making. Acting links metabolism with experience, habits, intent and many cognitive and linguistic techniques. While the mechanisms remain largely unknown, historically derived meanings connect up ways of acting, thinking and talking in an extended human ecology.

What goes on under the furrowed brow usually matters less that how people live in a partly collective world. Using simplex motifs such as bidirectional coupling, inter-bodily effects ground a person's cognitive and perceptual control. During a life history, agents become singular individuals whose sponsored ways are supplemented by skills in authoring actions. At times, people surprise themselves by using inhibition and a predictive brain to anticipate how others may behave. They track likely behaviour by using simulation to connect sense-saturated coordination with culture. In historical time, shared standards led to mastery of, among other things, tools, fire and language. By achieving reliable effects, people refine procedures. Physical and social affordances drive cultural and cognitive change. As first argued by Donald (1991), mimetic skills may suffice to explain why hominids diverged from other primates. Gradually, they came to see the potential of a resource (tool or object) as detachable from their own action. If this is correct, the principle of cognitive separation would itself favour increasing use of second-order resources.[8] Selection would favour practice, skill development and an ability to anticipate what to do now (and later). Using interactivity, modern individuals became masters of artifice. We even picture ourselves—or our fellows—as like Rodin's *penseur*. And that, we suggest, is truly extraordinary.

References

Adams, D. (1979). *The hitchhiker's guide to the galaxy*. London: Baker.

Auvray, M., Lenay, C., & Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, *27*, 32–47.

Baber, C. (2013). Distributed cognition at the crime scene. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Ball, L. J., & Litchfield, D. (2013). Interactivity and embodied cues in problem solving and insight: further contributions to a "theory of hints". In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Barbieri, M. (2009). A short history of biosemiotics. *Biosemiotics*, *2*, 221–245.

Ben-Naim, J., Bonnefon, J. F., Herzig, A., Leblois, S., & Lorini, E. (2013). Computer-mediated trust in self-interested expert recommendation. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Bernstein, N. A. (1967). *The co-ordination and regulation of movements*. Oxford: Pergamon Press.

Berthoz, A. (2010). La conscience du corps. In A. Berthoz & B. Andrieu (Eds.), *Le Corps en acte*. Nancy: Presses Universitaires de Nancy.

Berthoz, A. (2012). *On simplexity: simplifying principles for a complex world*. New Haven: Yale University Press.

Blair, G., & Cowley, S. J. (2003). Language in iterating activity: microcognition remembered. *Alternation*, *10*, 132–162.

[8]Cowley (2012) takes a more radical view: given mimesis, he argues that spoken language—like writing and computers—draws on the co-evolution of h. sapiens and the extended ecology (see also, Donald 2012).

Bottineau, D. (2010). Language and enaction. In J. Stewart, O. Gapenne, & E. Di Paolo (Eds.), *Enaction: toward a new paradigm for cognitive science* (pp. 267–306). Cambridge: MIT Press.

Bottineau, D. (2012). Remembering voice past: languaging as an embodied interactive cognitive technique. In *Proceedings of the humanities international forum*. Moscow: Russian State University for the Humanities.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton and Co.

Clark, A. (1996). Magic words: how language augments human computation. In P. Carruthers & J. Boucher (Eds.), *Language and thought: interdisciplinary themes* (pp. 162–183). Cambridge: Cambridge University Press.

Clark, A. (2008). *Supersizing the mind: embodiment, action and cognitive extension*. Oxford: Oxford University Press.

Cowley, S. J. (2003). Distributed cognition at three months: mother-infant dyads in kwaZulu Natal. *Alternation*, *10*, 229–257.

Cowley, S. J. (2004). Contextualizing bodies. *Language Sciences*, *26*, 273–299.

Cowley, S. J. (2007). Cognitive dynamics and distributed language. *Language Sciences*, *29*, 575–583.

Cowley, S. J. (2011). *Distributed language*. Amsterdam: Benjamins.

Cowley, S. J. (2012). Mimesis and language: a distributed view. *Interaction Studies*, *13*, 17–40.

Cowley, S. J., & Vallée-Tourangeau, F. (2010). Thinking in action. *AI & Society*, *25*, 469–475.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown & Company.

Dewey, J. (1896). On the reflex arc in psychology. *Psychological Review*, *3*, 357–370.

Donald, M. (1991). *The origins of the modern mind*. Cambridge: Harvard University Press.

Donald, M. (2001). *A mind so rare*. New York: Norton.

Donald, M. (2007). The slow process: a hypothetical cognitive adaptation for distributed cognitive networks. *Journal of Physiology (Paris)*, *101*, 214–222.

Donald, M. (2012). Commentary on Cowley's chapter. *Interaction Studies*, *13*, 41–49.

Evans, J. (2010). *Thinking twice: two minds in one brain*. Oxford: Oxford University Press.

Fioratou, E., & Cowley, S. J. (2009). Insightful thinking: cognitive dynamics and material artifacts. *Pragmatics & Cognition*, *17*, 549–572.

Froese, T., & Gallagher, S. (2012). Getting Interaction Theory (IT) together: integrating developmental, phenomenological, enactive, and dynamical approaches to social interaction. *Interaction Studies*, *13*(3), 436–468.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Giere, R. N. (2004). The problem of agency in scientific distributed cognitive systems. *Journal of Cognition and Culture*, *4*, 759–774.

Giere, R. N. (2011). Distributed cognition as human centered although not human bound: reply to Vaesen. *Social Epistemology: A Journal of Knowledge, Culture and Policy*, *25*, 393–399.

Gintis, H., & Gintis, H. (2011). Gene–culture coevolution and the nature of human sociality. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*, 878–888.

Hacking, I. (1999). *The social construction of what?* Cambridge: Harvard University Press.

Heidegger, M. (1971). Language. In *Poetry, language, thought* (pp. 185–208). London: Harper Collins.

Hemmingsen, B. (2013). The quick and the dead: on temporality and human agency. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, *7*, 174–196.

Hutchins, E. (1995). How a cockpit remembers its speed. *Cognitive Science*, *19*, 265–288.

James, W. (1890). *The principles of psychology* (Vol. I). New York: Dover.

Järvilehto, T. (2000). The theory of the organism-environment system: IV. The problem on mental activity and consciousness. *Integrative Physiological and Behavioral Science*, *35*, 35–57.

Järvilehto, T. (2009). The theory of the organism-environment system as a basis of experimental work in psychology. *Ecological Psychology*, *21*, 112–120.

Järvilehto, T., Nurkkala, V. M., & Koskela, K. (2009). The role of anticipation in reading. *Pragmatics & Cognition*, *17*(3), 509–526.

Jones, P. (2013). You want a piece of me? Paying your dues and getting your due in a distributed world. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Kiousis, S. (2002). Interactivity: a concept explication. *New Media & Society*, *4*, 355–383.

Kirsh, D. (1997). Interactivity and multimedia interfaces. *Instructional Science*, *25*, 79–96.

Kirsh, D. (2013). Thinking with external representations. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic actions. *Cognitive Science*, *18*, 513–549.

Kiverstein, J., & Farina, M. (2011). Embraining culture: leaky minds and spongy brains. *Teorema*, 32.2, 35–53.

Kravchenko, A. (2007). Essential properties of language: why language is not a digital code. *Language Sciences*, *29*, 621–650.

Kravchenko, A. (2011). How Humberto Maturana's biology of cognition can revive the language sciences. *Constructivist Foundations*, *6*, 352–362.

Linell, P. (2005). *The written language bias in linguistics: its nature, origins and transformations*. London: Routledge.

Love, N. (2004). Cognition and the language myth. *Language Sciences*, *26*, 525–544.

Lumsden, C. J., & Wilson, E. O. (1981). *Genes, mind, and culture: the coevolutionary process*. Cambridge: Harvard University Press.

Markoš, A., Švorcová, J., & Lhotský, J. (2013). Living as languaging: distributed knowledge in living beings. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Marr, D. (1982). *Vision: a computational investigation into the human representational system and the processing of visual information*. San Francisco: Freeman.

Maturana, H. R. (1975). The organization of the living: a theory of the living organization. *International Journal of Man-Machine Studies*, *7*, 313–332.

Mead, G. H. (1932). *The philosophy of the present*. Chicago: Open Court.

Monod, J. (1966). From enzymatic adaptation to allosteric transitions. *Science*, *154*, 475–483.

Neumann, M., & Cowley, S. J. (2013). Human agency and the resources of reason. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Pattee, H. H. (1996). The problem of observables in models of biological organizations. In E. L. Khalil & K. E. Boulding (Eds.), *Evolution, order, and complexity* (pp. 249–264). London: Routledge.

Pattee, H. H., & Rączaszek-Leonardi, J. (2012). *Laws, language and life: Howard Pattee's classic papers on the physics of symbols*. Dordrecht: Springer.

Peirce, C. S. (1940). *The philosophy of Peirce*. London: Routledge. J. Buchler (Ed.).

Perry, M. (2013). Socially distributed cognition in loosely coupled systems. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.

Rączaszek-Leonardi, J., & Cowley, S. J. (2012). The evolution of language as controlled collectivity. *Interaction Studies*, *13*, 1–16.

Ross, D. (2007). H. sapiens as ecologically special: what does language contribute? *Language Sciences*, *29*, 710–731.

Sharov, A. A. (2010). Functional information: towards synthesis of biosemiotics and cybernetics. *Entropy*, *12*, 1050–1070.

Spiekermann, K. (2013). Judgement aggregation and distributed thinking. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.
Spurrett, D., & Cowley, S. J. (2004). How to do things without words. *Language Sciences*, *26*, 443–466.
Steffensen, S. V. (2011). Beyond mind: an extended ecology of languaging. In S. J. Cowley (Ed.), *Distributed language* (pp. 185–210). Amsterdam: Benjamins.
Steffensen, S. V. (2013). Human interactivity: problem-solving, solution-probing and verbal patterns in the wild. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.
Stewart, J., Gapenne, O., & Di Paolo, E. (Eds.) (2010). *Enaction: toward a new paradigm for cognitive science* (pp. 33–87). Cambridge: MIT Press.
Stuart, S. A. (2010). Enkinaesthesia, biosemiotics and the ethiosphere. In S. J. Cowley, J. C. Major, S. V. Steffensen, & A. Dinis (Eds.), *Signifying bodies: biosemiosis, interaction and health* (pp. 305–330). Braga: Portuguese Catholic University Press.
Thibault, P. J. (2011). First-order languaging dynamics and second-order language: the distributed language view. *Ecological Psychology*, *23*, 210–245.
Thompson, E. (2007a). *Mind and life*. Cambridge: Harvard University Press.
Thompson, E. (2007b). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge: Belknap Press.
Vallée-Tourangeau, F., & Villejoubert, G. (2013). Naturalising problem solving. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: interactivity, computation and human artifice*. Dordrecht: Springer.
Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind: cognitive science and human experience*. Cambridge: MIT Press.
von Uexküll, J. (1926). *Theoretical biology*. New York: Harcourt, Brace & Co.
Wallot, S., & Van Orden, G. (2011). Grounding language in the anticipatory dynamics of the body. *Ecological Psychology*, *23*, 157–184.
Wheeler, M. (2004). Is language the ultimate artifact? *Language Sciences*, *26*, 693–715.
Wittgenstein, L. W. (1958). *Philosophical investigations* (2nd ed.). Oxford: Blackwell.

Author Index

A
Adams, D., 256, 270
Adams, F., 4, 9
Ahmadjian, V., 87, 92
Aitken, K. J., 20, 29
Aizawa, K., 4, 9
Alač, M., 149, 168
Alba, J. W., 195, 206, 221
Alberts, B., 74, 78, 90
Aliner, G., 220
Allis, C. D., 83, 84, 90
Anderson, A. H., 236
Anderson, M. L., 2, 9, 23, 27
Anfinsen, C. B., 77, 90
Åqvist, L., 61, 62, 68
Arrow, K. J., 32, 46, 50
Artman, H., 133, 145, 149, 168
Ashburner, J., 10
Atkinson, P., 153, 166, 168
Austin, J. L., 97, 110
Auvray, M., 5, 9, 268, 270
Aydede, M., 4, 10, 193, 199, 221, 238, 252

B
Baber, C., 6, 7, 9, 131, 138, 142, 143, 145, 146, 257, 264, 267, 270
Baddeley, A. D., 179, 193
Ball, L. J., 6, 7, 9, 207, 219, 223, 226, 232–234, 236, 238, 267, 270
Bandura, A., 101, 110
Bang, J. C., 208, 210, 213, 219
Bang, M., 136, 145
Barbieri, M., 75, 91, 269, 270
Bard, E. G., 231, 236
Baron, R. M., 198, 220
Baron-Cohen, S., 237
Barrett, E. C., 146
Barsalou, L. W., 4, 9, 229, 232, 235, 236
Bartlett, F. C., 108, 110
Baslev, T., 238
Bateson, M. C., 20, 27
Battersby, W. S., 228, 237
Bayliss, A. P., 231, 237
Becchio, C., 230, 236, 237
Bekkering, H., 239
Belpaeme, T., 3, 9
Ben-Naim, J., 6, 9, 26, 27, 53, 256, 270
Bender, M. B., 237
Bergman, A., 78, 91
Berkowitz, L., 206, 220
Bernsen, N. O., 98, 110
Bernstein, N. A., 266, 270
Bertalanffy, L., 90, 91
Berthoz, A., 259, 270
Bertone, C., 237
Bhattacharyya, A. K., 238
Bickhard, M., 17, 27
Bilalić, M., 251, 252
Black, D., 46, 50
Blackmore, S., 100, 110
Blair, G., 259, 270
Blanchette, I., 226, 237
Blandford, A., 133, 145, 146
Bonnefon, J.-F., 9, 27, 46, 50, 53, 270
Boshuizen, H. P. A., 239
Bottineau, D., 5, 9, 268, 271
Boudon, R., 25, 27
Bovens, L., 41, 42, 44, 48–50
Bradlow, E. T., 69
Bråten, S., 20, 27, 199, 220, 221
Brennan, S. E., 230, 231, 233, 236–239
Brooks, R., 230, 237
Brorsen, A.-K., 203, 219, 220
Bruner, J., 20, 27, 95, 110

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8, © Springer-Verlag London 2013

Buckley, B., 57, 69
Burningham, C., 239
Burt, R., 69
Busby, J. S., 133, 145, 226, 237
Butler, M., 138, 145, 146
Butterworth, G., 230, 237

C
Cacioppo, J. T., 57, 69
Calderwood, R., 146
Camerer, C., 69
Carlston, D. E., 57, 69
Carroll, S. B., 77, 89, 91
Castelfranchi, C., 61, 68, 69
Castiello, U., 237
Cedrone, C. C., 239
Chabris, C. F., 189, 193
Chalmers, D., 4, 9, 32, 50, 220
Chapman, P. R., 230, 237
Charlton, D., 146
Charness, N., 230, 237
Chellas, B. F., 62, 68
Chemero, A., 4, 5, 9, 200, 220
Chen, X., 237, 238
Chen, Y., 236
Chicoisne, G., 48, 51
Chomsky, N., 117, 259, 267, 271
Clark, A., 4, 9, 32, 47, 50, 106, 110, 134, 145, 174, 196, 197, 201–203, 218–221, 267, 271
Clinton-Cirocco, A., 146
Cohen, P. R., 61, 62, 68
Coleman, J., 25, 27
Conte, R., 61, 68
Corkum, V., 230, 237
Cowley, S. J., 1, 3, 5–10, 15, 17–21, 23, 27–29, 90–92, 111, 112, 129, 197–199, 218–221, 259, 261–265, 269–273
Craik, K., 22, 28
Crawford, T., 238
Cronin, B., 15, 28
Cross, J., 145
Crystal, D., 210, 220
Csikszentmihalyi, M., 108, 110
Culverhouse, P. F., 239

D
Dallob, P., 228, 237
Dalzel-Job, S., 236
Damasio, A., 107, 110
Davidson, D., 24, 28
Davidson, E. H., 89, 91
Davis, G., 237
Davis, J. H., 69
Davis, K., 28
Dawkins, R., 75, 91, 99, 110
De Jaegher, H., 5, 9
DeCasper, A. J., 28
Demolombe, R., 61, 69
Dennett, D., 4, 9, 23, 28, 259, 267, 271
Desai, V., 218, 220
Dewey, J., 259, 266, 271
Di Paolo, E., 5, 9, 10, 15, 28, 29, 237, 271, 273
Dickinson, C. A., 237, 238
Dietrich, F., 36, 37, 40, 50
Dirks, K. T., 69
Dollard, D., 206, 220
Dominowski, R., 228, 237
Donald, M., 199, 220, 259, 270, 271
Donovan, T., 238
Doob, L., 220
Døør, J., 208, 210, 213, 219
Douglas, A. E., 85, 86, 91
Doyle, A. C., 132, 133, 135, 145
Dreyfus, H. L., 138, 145
Dreyfus, S. E., 138, 145
Driver, J., 230, 237
Dror, I. E., 51, 133, 140, 142, 145, 146
Dryzek, J., 46, 50
Duchowski, A. T., 239
Dunbar, K., 226, 237
Duncker, K., 231–233, 237
Dunn, J. R., 58, 68
Durkheim, E., 14, 24, 28

E
Eastwood, J. D., 238
Eco, U., 73, 91
Eika, B., 238
Ekman, P., 111
Ellis, J., 151, 169
Emerson, E. A., 61, 69
Emery, N. J., 235, 237
Engestrom, Y., 148, 168
Epel, D., 77, 83, 91
Ericsson, K. A., 224, 237
Euden, G., 221, 253
Evans, J., 263, 264, 268, 271
Eymard, L. A., 252

F
Faltýnek, D., 76, 87, 90, 91
Farina, M., 259, 272
Feld, S. L., 50
Fenske, M. J., 237
Ferrin, D. L., 54, 69
Fester, R., 86, 91
Fields, R. E., 169

Fifer, W. P., 20, 28
Fioratou, E., 261, 271
Fischer, I., 57, 69
Fischer, M. H., 238
Fishkin, J., 51
Flin, R., 138, 146
Flom, R., 230, 236, 237
Flor, N. V., 133, 146
Forsythe, W., 176, 193
Fowler, C. A., 19, 28
Frackowiak, R. S. J., 10
France, L., 103, 110
Friesen, C. K., 238
Friesen, W. V., 111
Frischen, A., 237
Frith, C. D., 10
Froese, T., 262, 271
Furness, D., 133, 146
Furniss, D., 148, 168

G
Gadian, D. G., 10
Galinsky, A. D., 69
Gallagher, S., 262, 271
Galosia, M., 200, 202, 213, 220
Gapenne, O., 10, 29, 237, 271, 273
Garbis, C., 133, 145
Garfinkel, H., 167, 168
Geake, J. G., 251, 252
Gegenfurtner, A., 230, 237
Gehring, W., 88, 91
Georgeff, M. P., 61, 69
Gerjets, P., 238, 239
Gholson, B., 243, 252
Gibbs, R., 229, 235, 237
Gibson, J. J., 8, 10, 18, 28, 100, 111, 136, 146, 230, 237, 262, 266, 271
Giere, R. N., 148, 168, 243, 252, 258, 262, 264, 268, 271
Gilbert, S. F., 77, 83, 91
Gilboa, A., 111
Gilhooly, K. J., 228, 237, 238, 253
Gino, F., 58, 69
Gintis, H., 257, 271
Gobet, F., 252
Goffman, E., 19, 28
Goldman, A. I., 32, 45, 50
Good, C. D., 10
Good, M. C., 78, 91
Goodwin, C., 197, 220
Graham, L. E., 92
Gramopadhye, A. K., 238, 239
Grant, E. R., 232, 237
Grant, V., 146
Greenstein, J. S., 238, 239
Grofman, B., 40, 50
Grudin, J., 160, 168
Grygar, F., 91
Guthrie, L. G., 253

H
Hacking, I., 260, 265, 271
Hadjiilieva, K., 238
Hagoort, P., 20, 23, 30
Haider, H., 238, 252
Hajnal, L., 91
Hall, B. K., 80, 91
Halverson, C. A., 116, 126, 127, 129, 148, 149, 168
Hammersley, M., 153, 166, 168
Hanna, J. E., 230, 236, 237
Hansen, K., 103, 111
Hansen, P. C., 251, 252
Harel, D., 61, 69
Harel, I., 203, 220
Harnad, S., 3, 10, 23, 28, 51, 133, 145
Harris, R., 17, 23, 28, 93, 95–97, 99, 101, 102, 109, 111, 113, 116, 117, 119, 123, 124, 126, 129
Harvey, N., 57, 69
Hassabis, D., 102, 108, 111
Hauge, T. S., 20, 29
Havard, C., 236
Hazelhurst, B. L., 149, 168
Hearn, V., 221, 253
Hearst, E. S., 189, 193
Heath, C., 148, 156, 168
Hedstöm, P., 25, 28
Hegarty, M., 189, 193
Heidegger, M., 74, 76, 91, 259, 271
Hemmingsen, B., 6, 10, 93, 196, 220, 257, 264, 265, 267, 271
Henderson, J. T., 238
Heritage, J., 167, 168
Hermans, H. J. M., 108, 111
Hershey, J. C., 69
Herzig, A., 9, 27, 53, 63, 69, 270
Hind, S.-L., 146
Hintikka, J., 62, 69
Ho, M.-W., 72, 91
Hodges, B. H., 23, 28, 116, 122, 124, 129, 198, 220
Hoffmeister, M., 85, 91
Hofstadter, D. R., 75, 91
Holder, B. E., 149, 168
Hollan, J., 5, 10, 59, 69, 133, 146, 151, 167, 168, 197, 199, 202, 217, 220, 269, 271
Holmes, J. G., 69

Holmqvist, K., 238
Holyoak, K. J., 238
Horswell, J., 135, 146
Hubley, P., 20, 30
Hughes, J. A., 153, 169
Hummel, J. E., 238
Hunter, J., 145, 146
Hurley, S. L., 3, 10
Husserl, E., 95, 111
Hutchins, E., 5, 10, 98, 113, 114, 117–119, 121–123, 125–129, 133, 146, 148–151, 154–158, 165–169, 193, 220, 226

I
Ito, T. A., 57, 69

J
Jagodzinski, A. P., 239
James, W., 95, 107–109, 111, 263, 271
Jarodzka, H., 230, 231, 238, 239
Jarosz, D. F., 92
Jarret, N., 230, 237
Järvilehto, T., 7, 8, 10, 28, 98, 100, 106, 110, 111, 197, 199, 200, 220, 256–258, 260, 271, 272
Jefferson, G., 221
Jenuwein, T., 90
Johnson, A., 90
Johnson-George, C. E., 56, 69
Johnsrude, I. S., 10
Jones, P. E., 6, 10, 105, 111, 113–116, 118, 120, 122, 124, 126, 128, 129, 257, 261, 264, 272
Jørgensen, M. W., 95, 111

K
Kamhi, A. G., 252
Kaplan, C., 228, 238
Kauffman, S. A., 76, 91
Kidd, P., 237
Kingstone, A., 236, 238
Kintsch, W., 224, 237
Kiousis, S., 260, 272
Kirsh, D., 6, 10, 14, 173–175, 180, 184, 193, 198, 199, 220, 223–225, 227, 235, 236, 241, 242, 252, 257, 260, 261, 268, 269, 271, 272
Kiverstein, J., 259, 272
Klausen, T., 133, 146, 150, 169
Klein, G. A., 138, 146
Kleinke, C. L., 225, 229, 233, 235, 238
Kleisner, K., 76, 90, 91
Knoblich, G., 231, 238, 239, 243, 247–250, 252
Knudsen, G. M., 103, 111
Kodner, R., 92
Kokinov, B., 227, 238
Kornhauser, L. A., 33, 50
Koskela, K., 28, 272
Kounios, J., 228, 239
Kozen, D., 69
Kramer, R. M., 54, 69
Kratochvíl, Z., 91
Kravchenko, A., 17, 28, 265, 272
Krupinski, E. A., 230, 238

L
Landrum, R. E., 228, 238
Larkin, J. H., 225, 238
Latour, B., 241, 252
Lave, J., 226, 238, 243, 252
Leblois, S., 9, 27, 53, 270
Lee, K., 237, 239
Lehtinen, E., 237
Leijenhorst, H., 136, 146
Lenay, C., 9, 270
Leonardo, A. M., 239
Levenson, R. W., 107, 111
Levesque, H. J., 61, 62, 68
Levine, B., 111
Levine, J. M., 169
Levitt, M., 191, 193
Lewis, J., 90
Lewis, J. D., 56, 69
Lhotský, J., 10, 71, 272
Lim, W. A., 91
Lindheim, O., 124, 129
Lindquist, S., 78, 80, 92
Linell, P., 17, 23, 28, 87, 91, 97, 101, 111, 197, 209, 210, 221, 267, 272
List, C., 32, 33, 36–39, 41, 44, 46, 47, 50, 51
Litchfield, D., 6, 7, 9, 207, 219, 223, 231–234, 236, 238, 267, 270
Lleras, A., 228, 229, 233, 239
Lobmaier, J. S., 230, 231, 238
Locke, J., 17, 28
Logie, R. H., 179, 193
Løgstrup, K. E., 102, 109, 111
Lorenz, E. N., 208, 221
Lorini, E., 9, 27, 53, 61, 63, 69, 270
Loughlin, V., 7, 10
Love, N., 5, 10, 17, 18, 23, 28, 84, 91, 116, 124, 126, 129, 265, 272
Luchins, A. S., 242, 244, 245, 252
Luchins, E. H., 245, 252
Luff, P., 148, 156, 168
Lumsden, C. J., 257, 272
Luskin, R., 51

Lyon, P., 4, 10

M
MacDorman, K. F., 9, 19, 23, 28
Mackenzie, C. F., 146
Maddux, W. W., 54, 69
Madsen, P. M., 220
Maglio, P., 14, 28, 174, 181, 193, 268, 272
Maguire, E. A., 2, 10, 102, 108, 111
Maier, N. R. F., 195, 221, 228, 231, 238
Manning, D. J., 238
Margulis, L., 86, 91
Markoš, A., 6, 10, 71, 76, 84, 87, 89–91, 258, 263, 269, 272
Marks, J., 88, 91
Marr, D., 191, 193, 267, 272
Martin, W., 85, 91
Matsuda, R., 80, 91
Maturana, H. R., 5, 10, 19, 23, 24, 29, 265, 272
Maxwell, E., 237
Mayer, R. C., 54, 56, 69
McAllister, D. J., 56, 69
McGarry, B., 148, 169, 202, 217, 221
McLean, I., 51
McLeod, P., 252
McMaster, R., 145
Mead, G. H., 108, 111, 259, 272
Melser, D., 19, 29
Meltzoff, A. N., 230, 237
Merleau-Ponty, M., 95, 111
Middleton, D., 146, 148, 168, 169
Miller, N., 220
Moffatt, B., 243, 252
Monod, J., 3, 10, 75, 77, 92, 265, 272
Moore, C., 230, 237
Morgan, D., 252
Moscovitch, M., 111
Mowrer, O., 220
Mullen, E., 69
Müller, G., 91
Murphy, P., 228, 237
Murray, D., 156, 169

N
Nader, K., 108, 111
Nagel, E., 183, 193
Nalanagula, D., 230, 231, 233, 238
Nardi, B. A., 127, 129
Neider, M. B., 233, 237, 238
Nemeth, C., 136, 146
Neubauer, Z., 91
Neumann, M., 6, 10, 13, 263, 272
Newell, A., 151, 169, 224, 225, 238, 243, 253
Newman, J., 183, 193
Nicholson, H. B. M., 236
Nishimura, M., 239
Norman, D. A., 150, 169
Nurkkala, V. M., 28, 272
Nyström, M., 238

O
Ohlsson, S., 228, 231, 238, 247, 251–253
Olson, E. A., 142, 146
Ormerod, T. C., 136, 137, 146, 226, 236, 238
Overmann, J., 86, 92
Owen, G., 50

P
Paas, F., 239
Pablé, A., 97, 111
Paglieri, F., 61, 68
Panesar, S., 145
Paracer, S., 86, 92
Parsons, R., 239
Pattee, H. H., 15, 29, 257, 258, 262, 265, 272
Pauly, M., 37, 39, 51
Pearce, J., 241, 253
Pearson, R. D., 91
Pedersen, S. B., 196, 200, 218, 221
Pedone, R., 231, 238
Peirce, C. S., 8, 10, 259, 266, 272
Pender, Z., 146
Péron, A., 146
Perret, D. I., 238
Perry, M., 6, 133, 146, 147, 153, 169, 202, 221, 263, 272
Pettit, P., 33, 36–39, 46, 47, 50
Phillips, L., 95, 111
Piaget, J., 21, 24, 29
Pick, A. D., 230, 236, 237
Plasters, C., 146
Poe, M. T., 87, 92
Poirier, P., 48, 51
Pomplun, M., 231, 237, 239
Port, R., 17, 29
Proffitt, D. R., 236, 239
Puppe, C., 37, 50
Putnam, H., 134, 146
Pylyshyn, Z., 190, 193

Q
Quine, W. V. O., 178, 194

R
Rabinowicz, W., 41, 42, 44, 48–50
Rączaszek-Leonardi, J., 17, 29, 262, 269, 272
Raff, M., 90
Raguso, R., 16, 29
Randall, D., 169

Rao, A. S., 61, 69
Rappaport, R. A., 71–74, 89, 92
Reddy, M. J., 17, 29
Redecker, D., 86, 92
Reid, F. J. M., 226, 239
Reinberg, D., 90
Reingold, E. M., 237
Reisberg, D., 183, 193
Rempel, J. K., 56, 69
Resnick, L. B., 148, 169
Rhenius, D., 238, 252
Ricciardelli, P., 237
Richardson, M. D., 169
Richter, L., 238
Ristic, J., 238
Ritter, H., 239
Robbins, P., 4, 10, 193, 199, 221, 238, 252
Roberts, K., 90
Roberts, K. H., 220
Rogers, Y., 133, 134, 144, 146, 151, 169
Rosch, E., 4, 11, 273
Rosenbaum, R. S., 108, 111
Rosenblum, L. D., 19, 28
Ross, D., 16, 29, 112, 269, 272
Rousseau, M., 56, 69
Rutherford, S. L., 78, 80, 92

S
Sacks, H., 117, 129, 198, 221
Sacks, O., 103, 111, 186, 194
Sadasivian, D., 230, 239
Sagan, D., 86, 91
Sager, L. G., 33, 50
Säljö, R., 237
Sapp, J., 85, 87, 92
Saussure, F. de., 17, 196, 220
Savage-Rumbaugh, S. E., 21, 25, 29
Scheerer, M., 206, 221
Schegloff, E. A., 117, 221
Scheiter, K., 238, 239
Schmidt, H. G., 239
Schoorman, F. D., 69
Schraagen, J. M., 136, 146
Schwaninger, A., 238
Schweitzer, M. E., 58, 59, 68, 69
Scott, K. M., 238
Seagull, F. J., 136, 146
Searle, J. R., 4, 10
Searls, D. B., 84, 92
Sears, R., 220
Sebanz, N., 236, 239
Segal, E., 247, 253
Shah, P. P., 69
Shanker, S. G., 29
Shapiro, D., 169
Shapiro, J. A., 3, 10
Sharov, A. A., 258, 265, 272
Sheperd, S. V., 229, 235, 239
Siegal, M. L., 78, 91
Sigtryggsdóttir, T., 21, 29
Silverman, D., 117, 129
Simon, H. A., 148, 150, 151, 169, 224, 225, 228, 238, 239, 243, 253, 256
Simon, M. A., 90, 92
Sitkin, S., 69
Skinner, B. F., 110, 111
Skowronski, J. J., 57, 69
Slovic, P., 57, 69
Smilek, D., 238
Smith, J. E. K., 70, 131
Smith, P., 145, 239
Smith, P. A., 131, 146
Smith, R. W., 228, 239
Sniezek, J. A., 57, 58, 69
Sparrow, B., 19, 30, 99, 112
Spiekermann, K., 6, 26, 27, 29, 31, 40, 50, 256, 273
Spence, M., 20, 28
Spivey, M. J., 229, 232, 235, 237, 239
Spurrett, D., 3, 9, 123, 129, 262, 273
Squire, L. R., 107, 112
Stampe, D. M., 237
Steffensen, S. V., 6, 7, 9, 10, 15, 29, 71, 92, 94, 98, 111, 112, 195–200, 202, 205, 207, 208, 218–221, 273
Stein, R., 231, 233, 239
Stern, D. N., 20, 29
Stetsenko, A., 116, 129
Stewart, E., 146
Stewart, J., 4, 5, 9, 10, 23, 29, 237, 257, 270, 271, 273
Strauss, A., 111, 166, 169
Strawson, P. F., 178, 194
Stuart, S. A. J., 19, 29, 273
Suchman, L. A., 226, 239
Švorcová, J., 10, 71, 76, 84, 88, 91, 92, 272
Swap, W. C., 56, 69
Symons, L. A., 230, 239

T
Tabachneck-Schijf, H. J. M., 225, 239
Taipale, M., 78, 92
Taylor, P. J., 146
Taylor, T. J., 29
Teasley, S. D., 169
Teilhard de Chardin, P., 94, 112
Teuber, H. L., 237

Thibault, P. J., 15, 18, 29, 71, 92, 202, 210, 218, 221, 260, 265, 268, 273
Thomas, L. E., 228, 229, 233, 239
Thompson, E. T., 4, 10, 11, 16, 23, 29, 257, 273
Tiddeman, B. P., 238
Tillack, A. A., 238
Timpka, T., 136, 145
Tinbergen, N., 107, 112
Tipper, S. P., 237
Tiuryn, J., 69
Tomasello, M., 20, 21, 29, 235, 239
Tønsberg, G. H., 20, 29
Trevarthen, C., 20, 30, 199, 221
Tulving, E., 102, 107, 112
Turing, A. M., 3, 8, 10, 98

U
Ulbæk, I., 98, 110
Underwood, G., 230, 237

V
Vallée-Tourangeau, F., 1, 5–10, 27, 29, 73, 91, 92, 111, 129, 196, 219–221, 236, 241, 243, 245, 246, 253, 263, 264, 270–273
Van De Weil, M. W. J., 224, 239
Van Gog, T., 231, 239
van Hees, M., 37, 39, 51
Van Maanen, J., 166, 169
Van Orden, G., 9, 11, 263, 273
Van Swol, L. M., 57, 58, 69
Varela, F. J., 4, 11, 267, 273
Velichkovsky, B. M., 231, 233, 234, 239
Villejoubert, G., 6, 10, 241, 253, 261, 263, 273
Von Neumann, J., 190, 194
von Uexküll, J., 266, 273
Vygotsky, L. S., 32, 51, 189, 194

W
Waddington, C. H., 80, 92
Wærn, Y., 133, 145, 149, 168
Wallot, S., 9, 11, 263, 273
Walter, P., 90
Warshel, A., 191, 193
Watson, J., 242, 253
Wearing, D., 103, 104, 106, 112
Wegner, D. M., 99, 101, 105, 112
Weigert, A., 56, 69
Weisberg, R. W., 195, 206, 221
Weller, A., 249, 253
Wells, A., 3, 11
Wells, G. L., 142, 146
Wenger, D., 30
Wertheimer, M., 247, 253
Wertsch, J., 199, 221
Wheeler, M., 259, 267, 273
White, T. B., 58, 69
Willems, R., 20, 23, 30
Willis, M., 16, 29
Wilson, B. A., 103, 106, 112
Wilson, E. O., 257, 272
Wilson, M., 229, 236, 239
Wilson, R. A., 201, 202, 221
Wimsatt, W. C., 47, 51
Winerman, L., 112
Winocur, G., 111
Wittgenstein, L. W., 23, 30, 182, 194, 219, 221, 259, 273
Wright, P. C., 148, 160, 169
Wujec, L., 146

X
Xiao, Y., 146

Y
Yang, F., 145
Yaniv, I., 57, 70
Yates, J. F., 70
Yoveva, M., 238

Z
Zalatan, J. G., 91
Zanna, M. D., 69
Zelinsky, G. J., 237, 238
Zola, S. M., 107, 112
Zwaan, 229, 239

Subject Index

A
Action, 1–9, 14, 15, 17–24, 61–63, 93–96, 100–102, 113–116, 118–120, 124, 125, 127, 128, 143–145, 195–197, 255, 256, 258–266, 268–270
Action trajectory, 150, 258
Activity system, 165, 166, 168
Activity theory, 14, 116
Adaptivity, 5, 199
Adjacent possible, 27, 255
Advisor, 54–60, 63–66
Affective trust, 56, *see also* trust, computer-mediated trust
Affordance, 6, 7, 25, 183, 200, 202, 226, 242, *see also* material affordance, social affordance
Affordance landscape, 242
Agency, 6, 13–16, 18, 19, 21, 22, 24–27, 74–76, 93–101, 104, 105, 107–109, 113, 114, 116, 125, 127, 128, 145, 198, *see also* human agency, social agency, sociological agency
Agent, 6, 17, 22, 24, 26, 36, 42, 53, 56, 59, 61–68, 99, 100, 109, 136, 184, 251
Agentic trajectory, 107, 110
Aggregate (aggregated), 14, 16, 32, 33, 35, 36, 47, 54, 56, 60, 67, 257
Agreement in judgement, 23
Allele, 75
Alter-centric perception, 209, *see also* alter-centric perspective
Alter-centric perspective, 209–211, 213, 214, 217
Ambiguity, 120, 183
Anachrony, 94, 95, 104, 105
Analogy, 17, 18, 71, 84, 131, 134, 172, 181, 188
Anchor (anchoring), 97, 104–106, 174, 177, 178, 210, 250
Annotation, 176–178, 181, 211
Anonymity, 36, 39
Anterograde amnesia, 103, 104, 108
Anthropogenic, 4
Anticipation, 260, 264, 265
Anticipatory action, 258, 262
Anticipatory control, 257
Appraisal, 26, 124
Arrow Theorem, 32, 46
Artifact (artefact), 122, 136, 148, 216, 243, 255, 267
Artifactual (artefactual) reconfiguration, 207
Artificial Intelligence, 53, 66
Attention, 1, 2, 4, 7, 8, 21, 22, 25, 44, 100, 108, 109, 206–210, 223, 225, 226, 228–230, 232–236, 246, 247, 258, 259
Attentional guidance, 230, 232
Authorship, 93, 94, 100, 101, 105, 108, 110, 263, 269
Autobiographical memory, 24, 198, 199, 218
Autonomy, 15, 100, 159, 268
Awareness, 18, 24, 107, 158, 166, 167, 268–269

B
Behavioural mimicry, 54, 56
Being with others, 74
Benevolence, 54, 56–58, 60, 63–65
Bidirectional coupling, 1, 2, 8, 198, 259, 270, *see also* cognitive dynamics, coordination
Biohermeneutics, 72
Biological code, 74, 85
Biological function, 266
Biological syllogism, 74

S.J. Cowley, F. Vallée-Tourangeau (eds.), *Cognition Beyond the Brain*,
DOI 10.1007/978-1-4471-5125-8,

Biophysical skill, 26
Biosemiotics, 27, 72, 265
Biosphere, 71, 73–75, 85, 87
Bodily activity (movement), 1, 2, 214, 268
Bodily organization, 76
Body plan, 88
Boundary condition, 16, 26, 265
Broad mental content, 134

C
Causal relay, 258, 260, 262, 263, 267
Cellular dynamics, 266
Central dogma, 72, 74–76, 80, 87
Chaperon protein, 78
Chromatin, 83, 258, 263, 269
Chunk decomposition, 243, 247–249
Citation index, 67
Co-acting assemblage, 6, 73
Co-action, 19–23, 25, 94, 99
Code view, 17
Cognition in the wild, 46, 47, *see also* distributed cognition
Cognitive amplification, 191
Cognitive artifact, 189
Cognitive bias, 73
Cognitive control, 8, 269
Cognitive-cultural network, 6
Cognitive dynamics, 96, 195, 199, *see also* bidirectional coupling
Cognitive ecology, 195, 196, 200, 206, 263
Cognitive economy, 196
Cognitive ethnography, 5, 148, 161, 166, 226
Cognitive event, 199, 216, 218, *see also* event, trajectory analysis, pico scale
Cognitive event analysis, 195, 200, 201, 203, 218
Cognitive event trajectory, 200
Cognitive frame, 256
Cognitive powers (power), 23, 25, 85, 171, 175, 179, 217, 258
Cognitive process, 4, 59, 126, 148, 150, 151, 262
Cognitive processing system, 141
Cognitive properties of a person, 118
Cognitive psychology, 200
Cognitive resource, 2, 199, 211, 216
Cognitive skills, 2
Cognitive states, 133
Cognitive system, 6, 106, 150–152, 198, 201–203, 243, 262, *see also* distributed cognitive system
Cognitive terrain, 174, 176
Cognitive trajectory, 195, 202, 205, 214, 217
Cognitive trust, 56, 71
Cohabitation, 71, 86, 87
Collective activity, 145, 148
Collective control, 8, 256, 269
Collective intelligence, 45
Collective output, 35
Collective thinking, 31, 45
Communication, 5, 89, 96, 97, 113, 114, 116–119, 121, 123, 124, 128, 134, 151, 156, 160, 163, 218, 219, 226
Communicative activity, 117
Communicative cue, 235
Community of agents, 66
Competence, 39–41, 43, 44, 48, 49, 56–58, 60, 63–65, 67, 128, 263
Comprehension, 172, 173, 176, 229, 247
Computation, 2, 3, 7, 8, 59, 126, 149–151, 171, 172, 179, 180, 185
Computational architecture, 151, 164
Computational functionalism, 198
Computer mediated judgement, 55
Computer mediated trust, 6, 55, 59, 60, 66, *see also* affective trust, trust
Computer metaphor, 113, 128
Concerted action, 18
Conclusion based procedure, 37, 38, 41–45, 48–50
Condorcet Jury Theorem, 32, 40, 41
  competence assumption, 40
  independence assumption, 40
Conduit, 16, 17
Confidentiality, 58, 64
Consciousness, 4, 23, 35, 94, 99, 100, 104, 109, 110
Consensual domain, 5, 24, 265, 268
Constraint relaxation, 231, 243, 247–249
Construction, 14, 59, 74, 153–155, 157–161, 165, 171, 187, 188, 196, 198, 199, 202, 219, 229, 249, 263
Content, 2, 3, 108, 121, 134, 135, 150, 179, 184–186, 189, 260
Contextual factors, 140, 141
Contingency, 100, 162, 268
Conversation analysis, 198
Coordinate, 2, 18, 24, 128, 174
Coordination, 2, 8, 13–18, 22, 23, 25, 26, 59, 122, 126, 127, 153, 174, 195–197, 200, 212, 213, 216, 255, 256, *see also* cognitive dynamics, coordination, coupling
  sense-saturated, 2, 8, 13, 14, 195–197, 200, 213, 218, 255, 258, 260, 270, *see also* interactivity
Cost structure, 172, 173, 180, 185, 190

Coupling, 1, 2, 5, 6, 8, 14, 15, 24, 136, 174, 198, 251, 259, 265, 270, *see also* coordination, interactivity
Creative (creatively interpreted), 8, 108, 113, 124, 160, 217
Creativity, 7, 27, 94, 122, 127, 217, 218, 244
Crime scene examination, 131, 133–136, 138–144
Critical problem, 244, 246, 247
Cultural activity system, 226
Cultural-cognitive networks, 259
Cultural constraint, 19
Cultural determinism, 127
Cultural ecology, 98
Cultural learning, 24, 25
Cultural-organism systems, 74
Cultural organization, 6
Cultural patterns, 260
Cultural practice, 15, 126
Cultural psychology, 14
Cultural resource, 1, 2, 6, 13, 15, 16, 22, 24–26, 256, 265, 266
Cultural routine, 19
Culture, 2, 19, 20, 25, 71, 74, 88, 89, 113, 115, 117, 118, 123, 125–128, 163, 226, 227, 257, 267–270
Culture of the workplace, 163
Cybernetics, 90
Cycle duration, 160–162, 164
Cycle of information, 66

D

Decision maker, 53, 138
Decision making, 31, 47, 57, 181, 230
Decision process, 32
Decoding, 88, 124, 258
Deduction, 132, 133, 141, 144
Deixis
    alter-centric deixis, 211
    definition of deixis, 210
    deictic attention, 208
    deictic markers, 210
    deictic recalibration, 212
    deictic shift, 208–210
    deictic system, 210, 211
    first person deictic, 210, 211
    object deictic, 208
    personal deictic, 213, 217
    second person deictic, 213
    third person deictic, 209
    zero person deictic, 213
Deliberation, 46
Demand-led, 157, 263, 267
Democracy, 38
Dermal metaphysics, 199, 218
Design brief, 157, 166
Developmental pathway, 88
Diagnostic radiography, 223, 234
Dialogical system, 197, 202
Dialogicality, 94, 96
Dialogism, 209
Dictator, 36, 38, 39
Dictatorship, 38
Directionality, 95, 109, 149, 261
Discrepant awareness, 18, 24
Discursive dilemma, 33–35, 37–39, 46, 48
Distributed cognition, 6, 7, 32, 33, 46, 48, 93, 94, 96, 113, 114, 123, 125–128, 131, 133, 134, 143, 144, 147, 148, 150, 151, 196, 198, 202, 268, 269, *see also* cognitive system, systemic cognition
Distributed cognitive system (DCS), 106, 151, 152, 197, 199–203, 207, 208, 210, 212–214, 216, 217, 262
Distributed environments, 241
Distributed knowledge, 87
Distributed language, 48, 265
Distributed problem representation, 251
Distributed thinking, 6, 31–33, 35, 39, 45–48
Distributed thinking system, 35, 39, 45, 48
Distributed through time, 59, 202, 269
Division of labour, 151–153, 162, 163
Dramaturgy (dramaturgical), 94, 102, 106–109
Dynamic problem presentation, 251
Dynamical, 2–4, 17, 84, 207, 267, *see also* cognitive dynamics
Dynamicist stance, 106

E

Ecological stance, 199
Ecological system, 98
Ecology, 6, 8, 14–16, 77, 85, 93, 96, 98, 105, 195, 196, 199, 200, 206, 219, 263, 268–270, *see also* extended ecology
E-commerce, 67
Ecosystem, 78–80, 83, 84, 87, 88, 257
Edge of chaos, 9, 195, 217, 263
Education of attention, 8, 262
Effectiveness, 6, 22, 174
Effectivity, 3
Ego-centric perspective, 213
Einstellung, 244, 251
Embodied activity, 15
Embodied cognition, 229, 235
Embodied coordination, 14
Embodiment, 4, 8, 23, 255
Embryogenesis, 88
Emergence, 46–48, 93, 95, 201, 218

Emotional states, 216
Enabling condition, 207, 218
Enactivism, 257, 266
Enactivist paradigm, 4, 5
Enactment, 229
Encoding, 124, 171, 179, 182, 184–186
Encodingism, 17
Enkinaesthesia, 19, 259
Entification, 178
Environment, 6, 7, 14, 16, 25, 26, 46, 77, 78, 131–134, 142–144, 178, 198–200, 216, 217, 223–225, 241, 242, 245, 250, 251
Environmental constraint, 225
Epigenetic causation, 80
Epigenetics, 77, 80
Epistemic action, 14
Epistemic activity, 185
Epistemic conception of mind, 3, 4
Epistemic cut, 257, 265
Epistemological domain, 196
Epistemology, 31–33, 95, 99
Ethical judgement, 114, 128
Ethnography, 5, 14, 15, 148, 152, 161, 164–166, 226
Ethnomethodology, 198
Eusocial primate, 8
Evaluation, 65, 67, 107
Evaluative function, 227
Event, 7, 73, 133, 156, 158, 167, 195, 200–205, 207–209, 213, 216–218, *see also* cognitive event
Event identification, 200, 218
Event pivot, 201, 204, 205, 208, 209, 213, 216
Event trajectory, 200–202, 204, 205, 209, 216–218
Event trajectory analysis, 201, *see also* cognitive event analysis, pico scale
Evidence, 6, 7, 13, 17, 19, 26, 87, 96, 131–145, 189, 225, 226, 228–230, 234, 235, 251, 252, 260, 266, 267
Evolution, 72, 74–77, 79, 80, 85–89, 98, 150, 257, 258, 262
Evolution-development (evo-devo), 77
Expectation, 54, 67, 141
Experience, 2–5, 8, 9, 19–21, 23–25, 71–74, 82–85, 94, 95, 98–101, 104–110, 136–139, 143–145, 244, 245, 258, 259, 264, 265, 267–269
Experience of being, 72
Expert knowledge, 14, 54
Expert recommendation, 54, 59
Expertise, 7, 15, 54–57, 59, 138, 152, 178, 224, 225, 230, 235, 247
Explicit encoding, 171, 184
Extended cognitive system, 201, 243
extended system, 258, 260, 262, 263, 268
Extended ecology, 6, 8, 14–16, 199, *see also* ecology
Extended mind, 4, 5, 32, 133, 134, 185, 256, 259, 264
External computation, 179
External interaction, 187
External manipulability, 187
External representation, 171, 175, 182
External structure, 59, 73, 174–176, 178, 179, 182, 188, 193
Externalization, 176, 178, 181, 188
Eye contact, 207, 208, 216
Eye movements, 173, 223, 225, 229, 230, 232–236
Eye-tracking, 137, 139, 144, 233

F

Face time, 53, 54
Feedback, 67, 74, 75, 101, 134, 250, 261
Feeling tone, 20
Field force, 73
Finger marks (prints), 132, 136–138, 140–142, 144, 201
First-order activity, 5, 8, 21, 255, 265
First-order languaging, 210, *see also* languaging
First-person experience, 21, 265, 267
First personhood, 93, 94, 96, 99–102, 105–107, 109, 110
First-person phenomenology, 19
Fission, 25, 27
Fitness, 75
Fix cycle, 149, 155, 157, 161, 163, 164
Fixation bias, 205, 206, 208–210, 212–214, 217
Fixation position, 230
Fixed code, 84
Flexibility, 2, 3, 16, 22, 156, 160, 199
Flexible interpretation, 158
Flow, 8, 9, 22, 74, 75, 106, 108, 172, 196–199, 201, 205, 261, 265, 268
Formal language, 84, 85
Formal operation, 243
Formal scientific environment, 243
Formalization, 60, 61, 263, 267
Formulating, 200, 213, 260
Frame of reference, 1, 24, 114, 120, 125, 134, 177
Frustration, 205, 206, 216
Frustration-aggression hypothesis, 206
Functional information, 3, 8, 258, 263–267
Functional outcome, 260

Functional reorganization, 19, 261
Functional system, 151–155, 158, 160, 161, 164–167
Functional system of activity, 151
Functionalism, 2, 4, 198, 266
Functionalist, 4, 255–257, 266–268
Functionality, 5, 15, 83, 102, 107, 183, 262

G
Gaze, 1, 2, 7, 20, 21, 195, 207, 208, 210, 212, 214, 216, 223, 225, 230–235, 258
  coordination of, 216
Gaze following, 230, 231, 235
Genetic script, 75–77, 80, 84, 85, 87
Genome, 80, 87, 88, 90, 258
Genotype, 75
Gesture, 5, 18, 20, 172, 173, 178, 187, 229, 256, 262
Goal-directed action, 14
Goal-oriented behavior, 150
Goal state, 225, 227, 247
Granularity, 150, 165, 267
Grounded cognition, 7, 229, 235

H
Habits, 257–263, 266, 269
Heuristic, 2, 202, 227, 251
Heuristically-guided search, 225, 227
Highly rationalized environment, 148, 202, 217
Histone code, 84
Human action, 2, 8, 94, 100, 118, 167, 195–197, 218, 259, 261
Human agency, 6, 13–16, 18, 21, 24–27, 93, 94, 96, 97, 100, 101, 107, 264, *see also* agency, social agency, sociological agent, individual
Human artifice, 7, 9, 255, 256
Human autonomy, 268
Human beings, 15, 93, 95, 96, 99, 101, 102, 104, 105, 107, 115, 126, 196–199, 202, 216, 217
Human cognition, 5, 26, 85, 97, 126, 199, 217, 266, 267
Human cognitive agent, 262
Human conduct, 97
Human niche, 265
Human observing, 257
Human primate, 13, 14
Human rationality, 217
Human skills, 267
Human symbol grounding, 13, 18–20, 22, *see also* symbol grounding problem
Hunch, 250

Hypothesis generator, 134, 135, 232, 242
Hypothetical thinking, 209, 264

I
Ill structured problems, 148
Imagining structure, 192
Impasse, 7, 214, 227, 231, 247, 257
Implication, 132, 141, 161, 179, 191, 219
Impossibility result, 36, 37, 46
Impression formation, 57
In vitro, 226
In vivo, 226
Inconsistency, 38, 183
Individual, 15–17, 24–26, 31, 32, 34–49, 94, 96–99, 101–110, 116–118, 122–125, 127, 128, 134, 135, 150, 151, 225–227, 263–265, 268, 269, *see also* human agency
Induction, 132, 133, 144, 188
Information, 2, 3, 8, 23–25, 31–33, 60, 63–68, 73–75, 131–135, 142–145, 147–157, 159–161, 164–168, 179, 258–260, 263–267
Information flow, 74
Information pooling, 32, 40, 48
Information processing, 3, 53, 60, 79, 144, 151–153, 164–168, 233, 247, 258, 260
Inhibition, 9, 259, 264–266, 270
Input, 3, 4, 7, 8, 17, 36, 78, 94, 109, 141, 171, 218, 256–258, 260, 262, 266
Input-output sandwich, 3
Insight, 5, 7, 8, 150, 151, 185, 195, 197, 202, 207, 213, 214, 217, 228, 229, 231, 234, 247, 248, 268, 269
  insight problem, 7, 228, 229, 231, 247
  insight tasks, 223
Institution, 2, 7, 13–17, 24, 26, 47, 85, 217, 257, 265, 267, 268
Integration, 97, 103, 123, 136
Integrational linguistics, 95, 96, 113
Integrity, 54, 56, 58–60, 63–67, 181, 182
Intentionality, 7, 98, 102, 104, 198, 230, 264
Interaction, 5, 14, 46, 47, 53–55, 117, 148, 171–174, 176, 184, 188, 189, 191–193, 195, 196, 198, 199, 235, 236, 268, *see also* social interaction
Interactive cognition, 174
Interactive community, 87
Interactive cycle, 250
Interactivity, 1, 2, 5–9, 13–15, 18–20, 22–27, 195–201, 205–208, 210, 214, 216, 218, 219, 226, 250–252, 255, 256, 258, 260–265, 267, *see also* coordination, sense-saturated, coupling

Interactivity cycles, 260, 261
Interbodily activity, 6, 8
Intercellular ecology, 77
Interpersonal history, 54
Interpreting, 133, 162, 173, 188, 235, 260
Intersubjectivity, 20
Intuition, 138, 217, 218
Invoice case, 203
IQ, 250

J
Joining organ, 8, 260
Joint attention, 230, 236
Judgement aggregation, 8, 31–33, 35–37, 39, 41, 45–48, 256, 263, 264, 267
Judgement profile, 36

K
Knowledge, 14, 17, 18, 32, 39, 40, 53, 54, 56, 76, 90, 103–106, 136, 144, 152, 153, 209, 210, 224, 260–262
Knowledge acquisition, 224
Knowledge base, 53
Knowledge challenge, 32, 39, 40, 45
Knowledge distribution, 159
Knowledge repository, 226
Knowledge restructuring, 224

L
Language, 2, 4–8, 13–27, 71–73, 76, 84, 85, 87, 96, 97, 116–118, 123–125, 197, 205, 219, 259, 260, 265–270
Language myth, 17, *see also* myth of the language machine
Language stance, 15, 18, 22–26, 219
Language system, 17, 85
Languaging, 15–18, 22, 24–27, 71, 72, 76, 85, 90, 96, 210, 265, *see also* first order activity, first order languaging
Lineage, 16, 26, 83, 88, 89, 264
Linear biology, 87
Linguistic technique, 5, 268, 269
Literacy, 16, 18, 25
Lived experience, 4, 83, 261, 267–269
Lived world, 71
Living subject, 26
Living system, 200, 257
Local functional system, 160
Logical analysis, 135
Loosely coupled, 147, 148, 150, 153, 158, 160–167, 263
Loosely organized resources, 147
Low-cost categorisation, 219

M
Main cognizer, 202, 213
Majority, 34, 35, 37–40, 42, 43, 49, 50, 65, 152, 189, 229, 244, 246
Majority vote, 34, 37, 38, 40, 49
Majority voting, 34, 37, 40
Manipulability, 187, 249
Material affordance, 183, *see also* affordance, social affordance
Material symbol, 4, 5, 260, 267
Materiality, 174, 175, 182, 262
Materialization, 178
Meaning, 17, 18, 23, 24, 61, 71–74, 76, 88–90, 97, 98, 100–102, 104, 105, 120–124, 126, 218, 219, 256, 264, 265, 268, 269
Meaning making, 98, 265
Meaningful structure, 193
Mechanism, 4, 107, 180, 185, 219, 229, 232, 247, 258, 263, 269
Media, 18, 126, 148, 151, 154, 155, 159, 164, 172, 187, 259
Medical work, 149
Medium, 25, 71, 161, 179, 186, 187, 191, 198
Memory, 24, 54, 76, 88, 93, 97, 99, 102–110, 123, 150, 171, 172, 176, 178–181, 198, 199, 241
Mental projection, 173, 178
Mental puzzle, 245
Mental set, 7, 242, 244, 246, 251
Mental simulation, 181
Mental states, 4, 133, 135, 141, 145, 266
Mereological error, 17
Message, 67, 68, 83, 88, 89, 113, 119–124, 126
Metabolism, 16, 75, 85, 265, 269
Metaphor of life, 76, 90
Middle world, 99
Mimetic skill, 259, 270
Mind, 3–5, 8, 17, 25, 32, 54, 60, 104, 116–118, 128, 133, 134, 179–182, 229, 266, 268
Modal operator, 62
Modal simulation, 229
Modality, 62, 186
Moral responsibility, 113, 264
Moral sense, 122
Morals, 105
Motility, 257, 266
Motor simulations, 223, 229
Multi-agent application, 66
Multi-cellular individual, 87
Multi-level aggregation, 60
Multi-modal logic, 60
Myth of the language machine, 116, 117, *see also* language myth

N
Narrative, 95, 99, 108, 136, 137, 141–144, 209–211, 214, 217
Narrow mental content, 134
National Adult Reading Test, 250
Natural convention, 75
Natural encoding, 186
Natural laws, 198
Natural selection, 75
Negativity bias, 57
Network, 6, 66, 71, 78–80, 82–84, 86, 90, 125, 126, 175, 258, 267
Neurofunctional domain, 107
Niche construction, 196, 199, 219
Nine-dot problem, 195
Nine-month revolution, 20
Non-linguistic experience, 5
Non-local, *see also* virtual
 concerns, 261
 constraint, 258, 267, 269
 pattern, 266, 269
Non-situated elements, 218
Norm, 2, 21, 22
Normative world, 13, 73
Notational representation, 187
Nucleic acids, 74, 87

O
Object of thought, 171, 176, 178, 179, 182, 183
Objective idealism, 8
Objective judgement, 7
Office, 137–139, 154–156, 159, 160, 196, 204, 263
Ontogeny, 16, 76, 79, 82, 87, 259
Ontological substrate, 7, 195, 196
Ontology, 18, 95, 96, 98
Operation, 1, 2, 100, 126, 135, 149, 153, 162, 163, 173, 262
Organisational constraint, 7, 19
Organisational life, 148
Organisational structure, 152, 158, 159, 162, 163
Organisational systems analysis, 165
Organism-centrism, 196
Organism-environment system, 7, 106, 199, 200
Output, 3, 6, 7, 14, 17, 35, 36, 78, 94, 109, 218, 256–258, 260, 262, 265, 266, 269

P
Parity principle, 4, 203
Participatory sense-making, 5
Perçaction, 259, 260, 262, 263
Perception-action coupling, 136
Perceptual chunk, 248
Perceptual crossing, 5
Perceptual inquisition, 176
Perceptual subject, 208
Performance, 2, 7, 32, 33, 40, 41, 44, 63, 79, 80, 89, 118, 124, 160, 165, 202, 203, 234, 235, 250, 251
Performance of action, 124
Permanency, 241
Perseverance, 244, 245, 260
Persistence over time, 181
Persistent mutualism, 85
Personal responsibility, 113, 115
Phenomenological experience, 4, 27, 219
Phenomenology, 18, 19, 219, 257, 259
Phenotype, 74, 75, 80, 89
Physical space, 154, 180, 241, 242
Pico time-scale (Pico-scale), 7, 15, 197, 210, 215, 216, 219, 256, 263, 268, *see also* event trajectory analysis, cognitive event analysis
Population, 14, 75, 80, 88, 89, 263
Population-level regularity, 263
Pragmatism, 14
Predictive performance, 156
Predictors, 250
Premise-based procedure, 38, 41, 42
Presupposition, 73
Pretending, 19, 21
Priming, 223, 229, 230, 236
Principle of cognitive separability, 262
Principle of co-temporality, 96, 97
Principle of ecological assembly, 201, 218
Problem finding, 195, 196, 205, 263
Problem reframing point, 205, 207, 209
Problem solving, 2, 6, 7, 138, 147, 148, 150, 151, 153, 154, 156–163, 166–168, 172, 175, 200, 203, 206, 213, 214, 217, 218, 223–231, 233–236, 241–244, 246–252
Problem solving expertise, 152
Problem solving framework, 147
Problem solving strategy, 206
Problem space, 149, 158, 196, 200–202, 205, 206, 214, 225, 227, 243, 256
Procedure, 32, 37–39, 41–46, 48–50, 65, 117, 119, 122, 124, 125, 158, 202, 203, 207, 218, 243–246, 249
Process model, 251
Professional design practice, 226
Projection, 173, 178, 180, 193, 245
Protein assembly, 77
Protein community, 77
Protein ecosystem, 78, 79

Protein network, 78, 80, 82, 84
Protein players, 84
Protein-protein binding, 79
Protoconversation, 20
Pseudo-task structure, 201
Public key, 67, 68

R
Radical constructivism, 265
Radical embodied cognitive science, 5
Rational actors, 13, 14
Rationality, 14–16, 21, 22, 25, 26, 31–33, 35, 36, 38, 39, 46, 62, 217
  collective, 36
Reading aloud, 258
Real-pattern, 24, *see also* second order constraints
Real-time coordination, 22
Real-time flow, 196, 199
Rearrangement, 179–182, 187
Reasoning process, 48
Reciprocity, 26
Recognition-primed decision-making, 138, 140
Reductionism, 47, 199
  biological, 199
  computational, 199
Reformulation, 184–186, 187
Reification, 117, 176
Reiteration, 73, 261
Reliability
  positive and negative, 44
Replicable constraint, 262, 268
Replication, 75, 82, 207
Representation, 3–5, 7, 88, 108, 126, 148, 149, 151, 152, 164, 171, 172, 174–176, 182–184, 186, 187, 225, 242, 243, 247
Representational restructuring, 247
Representational transformations, 148, 165
Reproductive thinking, 251
Reputation, 57, 61, 66–68
Re-representation, 171
Resource, 2, 143, 147, 155, 156, 159, 160, 199, 211, 215–217, 242, 270
Resources-for-action, 134, 141, 143, 144
Resources of reason, 15, 22, 24, 26
Responsibility, 57, 60, 100, 113–115, 118, 122, 125, 128, 264
Restatement, 73, 184
Restructuring, 7, 224, 228, 231, 241, 242, 247
Result, 2, 6, 7, 13, 17, 25, 26, 36–38, 43, 45–47, 49, 59, 191, 197, 200–202, 209, 210, 218, 247
Retrograde amnesia, 104
Reversibility, 261
Ritual, 71, 72, 89, 149
Road-building, 153
Role reversal, 21
Running commentary, 139

S
Saccade, 233, 234
Safety-related rules, 154
Said, 14, 15, 17–20, 26, 90, 97, 101, 114, 131, 143, 189, 202, 210, 214
Satisficing, 226
Savant syndrome, 189
Scaffolding, 82, 161, 202, 223
Schema, 140, 246, 247, 251
Second-order constraints, 16, *see also* real pattern
Second-order cultural construct, 5, 18
Second-order language, 210
Second-order phenomena, 2
Second-order structure, 255
Secondary event pivot, 209
Segregational philosophy, 123
Self-certifying, 188
Self-directed behavior, 258
Self-motility, 258
Self-organisation, 15, 16, 158, 166
Self-organising, 3, 266, 269
Semiosis, 100, 108
Semiotic relations, 259
Sense-making, 5, 22–24, 74, 95, 96, 98–100, 102, 107, 109, 197, 258–260, 262, 264, 269
Sense of self, 24, 101, 105, 106, 257
Sense of selfhood, 108
Sense saturation, 196, 197, 200, 210, 212, 213
Sensorimotor action, 7
Sensorimotor activity, 5
Sensorimotor control, 7
Sensorimotor experience, 3
Sequentiality, 261
Set problem, 245, 246
Shared attention, 22, 208, 209, 216
Shared standards, 270
Signal detection theory, 235
Signhood, 126
Simplexity, 259
Simulation, 7, 181, 186, 189–192, 212, 216, 219, 232, 255, 270
Situated activity systems, 197
Situation awareness, 158, 166, 167
Skill development, 270
Social actor, 16, 22, 25
Social actor theory, 25

Social affordances, 13, 22, 26, 269, 270, *see also* affordance, material affordance
Social agency, 104, *see also* agency, human agency, sociological agency
Social choice theory, 31–33, 46
Social epistemology, 31, 32
Social interaction, 53–55, 117, 128, 148, 196, 198, 219, *see also* interaction
Social meshwork, 71, 267
Social normativity, 199
Social reality, 104
Social rules, 198
Social system, 89, 148
Socially embodied activity, 151
Sociocognitive science (view), 7, 13, 26, 255–257, 266, 268
Sociocultural environment, 14
Sociocultural patterns, 5
Sociocultural resources, 198, 218
Sociocultural system, 118
Socioculture, 197
Sociological agency, 24, 25
Sociology of science, 17, 252
Sociotechnical system, 147
Solution probing, 9, 205, 206, 213, 214, 217, 257
Spatial resolution, 251
Sponsorship, 93, 94, 100, 101, 105, 108–110, 264
Standard operating procedure, 158
Static environments, 241
Strange duality, 7, 74, 264, 267
Structural coupling, 5, 24, 265
Swarm intelligence, 45
Symbiogenesis, 77, 86
Symbiosis, 77, 85–87
Symbol grounding problem, 23, *see also* human symbol grounding
Symbol processing, 13, 23, 164
Synchronicity, 101
Synchronization, 174
Systematicity, 36, 37, 39
Systemic cognition, 6, 251, *see also* distributed cognitive system
Systemic function, 258
Systemic output, 14, 262, 269
Systemic psychology, 200
Systems level, 151
Systems theory, 90, 200

T
Task analysis, 165
Task-based model, 264
Team, 34, 114, 118, 128, 149, 153–159, 189, 226
Team-based game playing, 149
Team effort, 118
Technology, 7, 16, 17, 151, 153, 189, 251, 258
Temporal dynamics, 95, 106, 182, 184
Temporal integration, 103
Temporal operator, 62
Temporal pattern, 200
Temporal scale, 27
Temporality, 93–98, 102, 106, 109
Temporary works, 153, 154, 157, 160, 161, 166
Theory of hints, 223, 225, 227, 235
Theory of mind, 262
Thinking, 1–4, 6–9, 14, 15, 31–33, 35, 39, 45–48, 113, 114, 124–128, 171–174, 185–189, 241–245, 251, 252, 255–260, 262–270
Thinking in action, 1, 6
Thought, 3, 4, 8, 24, 108, 109, 115, 116, 124, 171, 176, 178, 179, 182–184, 186, 187, 192, 193, 234, 236, 244, 245
Thought process, 234
Time-scale, 2, 7, 15, 17, 27, 93–98, 107, 161, 162, 199, 255, 256, 259, 261, 263–265, 267, 268
Tool, 88, 137, 142, 171, 175, 214, 269, 270
Transactive memory, 99, 105
Transcription, 75, 77, 209, 258
Transfer, 7, 76, 78, 224, 230
Transformation, 126, 151, 164, 242, 243, 247
Transformation problem, 242, 243, 247
Transient extended cognitive system (TECS), 201, 202
Transient mental presence, 182
Transition point, 201, 204, 205, 207, 209
Translation, 4, 75, 77, 182, 188
Transmissibility, 82
Trust, 6, 8, 26, 53–61, 64, 66–68, 256, 263, 264, *see also* affective trust, computer mediated trust
Trustworthiness, 54, 56–60, 64, 68
Truth tracking, 40, 41

U
Ukuhlonipha, 20, 25
Ultimate artifact, 267
Umwelt, 15, 72, 89, 102
Unanimity, 65
Unidimensional alignment, 47
Universal domain, 36, 39
Universal domain axiom, 46
Unpredictability, 87, 160, 191

Unquiet, 107
Utility calculation, 16

V
Variation, 263, 265
Verbal pattern, 206
Vested interests, 53, 54, 57, 59, 66, 67
Virtual construct, 23, *see also* non-local
Visual complexity, 179
Visuo-spatial ability, 250
Vocal gestures, 216
Volition, 109
Von Neumann machine, 2, 3, 260

W
Water jar task, 242, 244, 249, 251
Well-defined boundaries, 147
Wording, 23, 75, 80, 121, 207, 208, 212, 214, 217, 219
Work activity, 152
Written language bias, 17, 87, 267

Z
Zeitgeber, 210, 216, 219