History, Philosophy and Theory of the Life Sciences

C.U.M. Smith
Harry Whitaker *Editors*

Brain, Mind and Consciousness in the History of Neuroscience

Springer

Brain, Mind and Consciousness in the History of Neuroscience

History, Philosophy and Theory of the Life Sciences

Volume 6

For further volumes:
http://www.springer.com/series/8916

C.U.M. Smith • Harry Whitaker
Editors

Brain, Mind and Consciousness in the History of Neuroscience

Springer

Editors
C.U.M. Smith (1930–2013)
Department of Vision Sciences
Aston University
Birmingham, UK

Harry Whitaker
Department of Psychology
Northern Michigan University
Marquette, MI, USA

ISSN 2211-1948 ISSN 2211-1956 (electronic)
ISBN 978-94-017-8773-4 ISBN 978-94-017-8774-1 (eBook)
DOI 10.1007/978-94-017-8774-1
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2014937790

Printed on acid-free paper

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

On January 4, 2013, Chris Smith passed away. This book represents an idea that he broached to me three years ago; his consummate scholarship and unflagging collegiality is reflected in both the scope and quality of the work herein. Robert Burns mused "O wad some Power the giftie gie us To see oursels as ithers see us!" Well, Chris, we saw you as an academic and a gentleman; thank you for making a part of our lives so special.

Whit

Acknowledgment

Without the unflagging assistance of Jenny Smith, it would not have been possible to complete this volume of essays. Je te remercie de tout coeur.

Whit

Contents

Introduction

This volume of essays grew from a workshop which we organized at the combined ISHN/Cheiron meeting at Calgary/Banff, Alberta, Canada in June 2011.[1] What did we hope to achieve? It seemed to us that the joint meeting of our two societies provided an opportunity to use the assembled expertise in the history of social sciences and neuroscience to examine once again that thorniest of all issues at the interface between science and the humanities: the problem of mind, or, to use David Chalmers' well-known phrase, 'the hard problem'.[2]

Let us begin by defining, so far as possible, how the 'hard problem' appears to us in our unchosen place at the beginning of the twenty-first century. Although there are many approaches, one of the best is provided by the so-called explanatory gap. Imagine a neuroscientist in a utopian future examining the auditory cortex. Imagine further that this all-knowing neurophysiologist is congenitally deaf. Imagine finally that he is examining the auditory cortex of a musician listening to a Haydn symphony. This neurophysiologist can see the fluxes of ions moving across the membranes of populations of neurons, the associated alterations in magnetic fields and the spurts of neurotransmitters diffusing across a myriad of synaptic junctions; from that data he will be able to reconstruct the acoustic patterns produced by the instruments of the orchestra. However, we suspect that our future neuroscientist will have no idea of what the musician is experiencing nor what caused the storm of applause at the end of the symphony. Similarly, of course, vice versa. The musician need have absolutely no knowledge, just as Haydn himself had no knowledge, of neuroscience, to experience the symphony. This is the so-called explanatory gap. How is the 'lived-through' experience related to the goings-on in the cerebral cortex? This is the 'hard problem'.

The paradigmatic instance of 'consciousness', the instance which lies at the core of the hard problem, is thus the 'raw feel', or 'quale'. It is what we live through when the dentist drills an aching tooth without benefit of anesthetic; it is what occurs

[1] ISHN = International Society for the History of the Neurosciences; Cheiron = The International Society for the History of Behavioural and Social Sciences.

[2] Chalmers (1995, 1996).

when we regard with pleasure a painting of Breughel the Elder. It is also what the two month baby experiences when he is given his bottle or when a pin penetrates his skin. It is what the puppy has when master carelessly steps on its tail, or the kitten enjoys when its fur is stroked. How is this related to fluxes of ions and chemicals in the grey matter of brains?

In recent years neuroscientists have intensively studied the cerebral activity correlated with conscious experience. There have been innumerable attempts to hone in on the so-called neural correlates of consciousness (NCCs). Patterns of cell activity have been detected, for instance, when rats thread their way through a maze and it has been shown that these same patterns recur in REM [dreaming] sleep; possibly the salient events of waking life are replayed in the rat's slumbers.[3] But, as ever, we are watching the neurophysiological correlates of consciousness; we are no nearer a handle on answering the hard problem.

There have, of course, also been huge developments in neuroimaging. Pictures derived from PET and fMRI, for example, are commonly described as showing the mind in action. It is said that we can see where mental arithmetic occurs, where words are formed, where lies are generated, even where our religious sentiments are located. What, of course, we are actually seeing is a computer generated image of, for example, changes in the levels of oxygen in the blood during these mental activities. Furthermore, a more comprehensive analysis shows that what is usually discussed in fMRI or PET based imaging studies are the 'hot spots' – activity above a cut-off threshold; typically large areas of the brain are activated at lower levels and, even more surprising, there are 'low spots', too, areas in which a suppression of activity is reliably associated with the variables of the experiment.[4] In other cases real time activity of the brain is recorded by EEG or MEG machines. But in these cases temporal resolution is accompanied by less reliable localization. One might suggest that there is at the present time a sort of neuroimaging analogy to Heisenberg's uncertainty principle: as with a fundamental particle, so in neuroimaging. If we can get a handle on *where* in the brain the event(s) are occurring (PET, fMRI) we have only an imprecise idea of *when* they are occurring; if we can determine *when* they are occurring (EEG, MEG) we are uncertain exactly *where*.[5] The solution is, of course, to combine both techniques. But whatever the level of sophistication and whatever the technical advances we are still only seeing the brain activity correlated with consciousness. We are not addressing the 'hard problem'. Furthermore, how is it that different regions of the cerebral cortex support quite different types of sensory consciousness? Various specific areas in the occipital lobe 'light up' (in fMRI or MEG scans) when the subject reports visual experience; quite different areas, this time in the temporal lobe, are activated when the subject reports an auditory experience.

---

[3] Ji and Wilson 2007.

[4] Whitaker and Hochman 1995.

[5] The analogy is, of course, not exact. There is no theoretical reason (as there is with the Heisenberg uncertainty relations) which prevents us localising consciousness-related brain activity simultaneously in both space and time. The limitation is only technological and may indeed be solved as instrumentation improves into the twenty-first century.

Yet, if these areas are examined at a fundamental level, indistinguishable fluxes of ions across lipoprotein neural membranes are to be found. Moreover, clever ‘rewiring’ experiments whereby retinal neurons are re-routed to the auditory cortex show that this cortex can also provide visual experience.[6] The re-wiring experiments create what might be called the Lord Edgar Douglas Adrian paradox. As is well-known, Adrian argued that since action potentials are the same in different nerves, it must be the brain area itself that determines the “quality” of a sensory experience. From the re-wiring experiments, it would appear that this in fact is not the case, hence “Adrian’s Paradox”. How is it that the same underlying biophysics and molecular anatomy correlates with such categorically different qualia as hearing and vision?

The problem also obtrudes on the output side. The words appearing on the computer screen as we write this introduction are not instances of automatic writing; we believe that we are responsible for what appears. Twenty-first century neuroscience has, however, cast doubt on the origin in consciousness of what seems to be voluntary, intentional, movements. Activity can be detected milliseconds, occasionally even seconds, before the conscious decision is made.[7] That we are conscious when we perform a voluntary act is simply a matter of definition.

Nevertheless, the vast onrush of neuroscientific research often completely bypasses any concern with the ‘hard problem’. From molecule and ion channel to cerebral cortex, neuroscientists report as though they find no significant hiatus. There are still problems aplenty and new techniques are all the time invented to solve them: but they are all ‘easy’ or, better, ‘tough’, problems in the sense defined above. Although we are still very far from closure we can have no serious doubt that the brain is a material or physical system from top to bottom. There are no apparent gaps into which we are required to insert any immaterial principle. Neuroscience is increasingly sophisticated, increasingly fascinating. But in a sense this sophistication, accompanied by an enormous amount of research, obscures, rather than addresses, the ‘hard problem’. It is all too easy to accept the neuroscientist’s statement that his techniques show the ‘mind’ at work, that soon we shall be able to read-off the thoughts of experimental subjects and look into their minds. It has to be remembered that these statements are a mere shorthand and that what is in fact being said is that we shall soon be able to follow the neural correlates of these subjective occurrences – not the occurrences themselves. As one of the great neurologists of the nineteenth century, John Hughlings Jackson, never tired of saying: it is a profound mistake to take the brain to be a ‘solid mind’.[8]

This volume of essays investigates some of the milestones along the road to how we arrived at this position. The essays look at how mind and brain have been assessed at different periods in the history of Western neuroscience. We start in classical antiquity and sample how the problem has appeared to neuroscientists (in the widest possible sense of that appellation) through to contemporary times. In a single volume

[6] Sharma et al. 1999.

[7] See Soon et al. 2008.

[8] Jackson 1874.

we cannot hope to provide a systematic review. All we hope to do is to bore down into the subject at different moments in its history. Has there been any progress in our understanding? Is the problem, as some have said, merely badly posed? Or is it, perhaps, simply insoluble?

Department of Vision Sciences
Aston University, Birmingham, UK
Christopher Smith
(Author was deceased at the time of publication)

Department of Psychology
Northern Michigan University
Marquette, MI, USA
Harry Whitaker

References

Chalmers DJ (1995) Facing up to the problem of consciousness. J Conscious Stud 2:200–219
Chalmers DJ (1996) The conscious mind. Oxford University Press, Oxford
Jackson JH (1874) On the scientific and empirical investigation of epilepsies. Med Press Circular 18–22; In: Taylor J (ed) (1931) Selected writings of John Hughlings Jackson. Hodder & Stoughton, London, p 170
Ji D, Wilson MA (2007) Coordinated memory replay in the visual cortex and hippocampus during sleep. Nat Neurosci 10:100–107
Sharma J, Angelucci A, Sur M (2000) Induction of visual orientation modules in auditory cortex. Nature 404:841–847
Soon CS, Brass M, Heinze HJ, Haynes JD (2008) Unconscious determinants of free decisions in the human brain. Nature Neursci 11:543–545
Whitaker HA, Hochman D (1995) Alternative interpretations of PET measurements. Hum Brain Mapp (Suppl 1):210

Chapter 1
Beginnings: Ventricular Psychology

C.U.M. Smith

> . . . this part which must be such as to control and govern the passage of the *pneuma* . . . is not the pineal body but the *epiphysis* that is very like a worm and is extended along the whole canal (Galen: *De usu partium corporis humani* (Transl. M.T. May))

Alfred North Whitehead famously said that all Western thought is but a series of footnotes to Plato. So let us start with Plato and, in particular, with that curious Pythagorean tract he published as a sort of addendum to the *Republic*. The *Timaeus* was in fact the only one of Plato's works known to the Latin West until the translation surge of original Greek texts after the fall of Constantinople in 1453 ushered in the Renaissance. It describes in a dreamy manner a primitive neurophysiology to set alongside the poetical psychology of the *Republic* and the *Phaedrus*. Thus in the *Timaeus* we find a tripartite neurophysiology to parallel the tripartite sociology of the *Republic*: the proletariat, the warriors and the philosophers being represented in the body by a concupiscent soul beneath the diaphragm, a warrior soul in the thorax, and an intellectual soul in the head. It is only the latter soul that has the prospect of immortality, returning in a Buddhist-like cycle, through the forms of lesser animals, depending on its performance during life. The two lower souls are not offered this opportunity and perish when the body dies. The point being made here is that from remote antiquity (and Plato's 'likely story' is derived from yet earlier Pythagorean thought) a distinction was made between an immortal and a mortal 'soul'.

Plato's greatest pupil, Aristotle, was also clear that the intellect was something different from the other forces that motivated the human body. He also divided these indwelling forces or souls into three kinds: vegetable, animal and rational. Although he does not follow Plato in assigning immortality to the rational soul, he nevertheless points out that it cannot be blended with the body, else it would acquire one or other

---

Author was deceased at the time of publication.

C.U.M. Smith (✉)
Emeritus, Department of Vision Sciences, Aston University, Aston Triangle,
B4 7ET Birmingham, UK
e-mail: a.m.jennysmith@gmail.com

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_1,

of the body's characteristics, (warmth, cold etc.) and this is not the case. In addition, whilst sensations fade with repetition (habituate, we would say), thought only becomes clearer with practice (*De anima 429b1*).

When we come to the Alexandrian neurophysiologists, Herophilus and Erasistratus, or what fragments of their work that have come down to us, we do not find this philosophical concern with the immortality or otherwise of the ratiocinative soul. Instead we find a more focused approach on the body's physiology. Nevertheless, a tripartite classification still obtains. Blood, nutritive spirit, is contained in the veins, vital spirit or *pneuma zotikon* in the arteries, and animal spirit or *pneuma psychikon* in the nerves. These vessels, intertwined down to and below the level of ordinary visibility, form the substance of the flesh.

The last great alumnus of the Alexandrian medical school, Claudius Galen, half a millennium after Herophilus and Erasistratus, developed the Alexandrian neurophysiology by showing experimentally that the arteries did not contain *pneuma zotikon* but, as William Harvey, was later to put it, nothing but the 'equivocal gore'. He also dissected the brain to show that it contained four ventricles—two anterior ventricles (our lateral ventricles), a middle ventricle (our third) and a posterior ventricle (our fourth). These cavities he observed to be full of spirit which, he argued, had been distilled from the blood. This spirit, *pneuma psychikon* or animal spirit, was destined to play a long and significant role in neuropsychology.

Galen also described two other structures which were to have similarly long and significant roles in medieval neuropsychology. One was the pineal gland or *conarium*, the other was a worm-like structure that, he says, extends along the wall of the whole passage between the middle and posterior ventricle.

Before his time the pineal was thought, he says, to have the same function as the pylorus of the stomach. Just as the pyloric valve controls the flow of nutrients from the stomach to the small intestine, so the pineal was thought to control the flow of *pneuma* from the middle to the posterior ventricle.[1] Galen will have none of this. He is clear that the pineal projects upward from the roof of the mid-brain and not downwards into the aqueduct as it would if it were to act as a valve. 'Since this gland [the pineal] is', he writes, 'attached not to the inside but to the outside of the ventricle, how could it, having no motion of its own, have so great an effect on the canal? . . . Why need I mention how ignorant and stupid those opinions are?'[2] In spite of Galen's vehemence we shall see that the pineal made a remarkable come-back in late medieval and Renaissance neuropsychology.

Nevertheless, Galen felt the need of *some* ventricular structure to control the flow of *pneuma* from the anterior to the posterior ventricles. This valving action was, he believed, undertaken by the 'worm-like' structure that he had found to line the passage between the two ventricles. 'Those versed in anatomy' he writes, 'have named it for its shape alone and call it the vermiform epiphysis [*vermis superior cerebelli*].'[3] It is

[1] Galen (trans. Brock AJ1916).

[2] Galen, *De usu partium corporis humani*, 1, 491

[3] Galen, *De usu partium corporis humani*, 1, 491

this, he writes, that regulates the flow of animal spirit from the middle to the posterior ventricle. We touch later on the problem of where and how and when Galen's epiphysis became confused with the pineal or conarium.

In addition to being an anatomist Galen was also an experimentalist. A famous piece of serendipity resulted in his discovery of one of the functions of the recurrent laryngeal nerve when he accidentally cut it and found that his experimental animal—in this case a pig—stopped squealing. Evidently the brain controlled the voice, that most expressive signature of the soul. He also investigated the significance of the ventricles. In his short work on breathing he remarks that:

> Since the emptying of the pneuma from the hollows of the brain, when it is wounded, at once makes men motionless and without sense, it must be surely that this pneuma is either the very substance of the soul or its primary organ.[4]

This uncertainty over whether the *pneuma psychikon* or animal spirit is 'the very substance of the soul or its primary organ' persists, as we shall see, for a millennium. Galen next proceeded to investigate the significance of the different ventricles. Pressing down on each in turn he showed that the posterior, our fourth, was the most important. Its destruction, he found, invariably led to the death of the animal. The importance of this ventricle was also emphasized by his observation that the cranial nerves originated from the medulla or from the floor of the fourth ventricle beneath the 'parencephalon', our cerebellum. He noted the pulsations of the brain when the skull had been removed and concluded that the brain, like the heart, acted as a pump to force fluid, in this case the *pneuma psychikon* within the ventricles, to flow out along the nerves. The significance of the vermiform valve is now apparent. This, of course, was long before Harvey established the true rate at which blood flows around the circulatory system. For Galen and the ancients the flow was only sluggish, so the analogy with the imagined flow of animal spirit was not far-fetched.[5]

Finally, and most importantly, Galen found that opening the ventricles during vivisection did not necessarily lead to the death of an animal. If the incision was not too severe and the wound healed, the animal regained its normal behavior. This seemed to him to answer the uncertainty he had expressed in his treatise on respiration. He concluded from this experiment that the animal spirit should not be confused with the soul, or principle of life. If it were, its escape would leave the animal dead. Rather, he reasoned, the *pneuma psychikon* was the soul's instrument. The soul was to be found elsewhere, most likely in the substance of the brain itself, not in its ventricles.

Galen did not assign different offices to the ventricles as did later authors. Although there were probably several predecessors, the first fully developed

---

[4] Galen on respiration translators. Furley and Wilkie 1984, p. 121.

[5] Galen is, in fact, somewhat undecided about how the *pneuma psychikon* makes its effects felt. In *De locis affectis*, book 1, chapter 7 (Siegel 1968, pp. 31–32), for example, he eschews the hydrodynamic model, and makes a comparison with the sudden strike of a ray of sunlight. This analogy resonates down the millennia and may be found in Islamic and Medieval texts and makes an appearance as late as the seventeenth century in Willis's *Cerebri anatome* (Willis 1681, p. 127).

ventricular psychology that has come down to us is due to Nemesius, the late fourth-century bishop of Emesa, a city in western Syria now known as Homs. It may be that medical observations between the time of Galen in the second century CE and those of Nemesius in the fourth played a role in this development.[6] However this may be, Nemesius' text initiates the great tradition of ventricular psychology, lasting well over a millennium, in which the brain was supposed to contain three hollow spaces, or cells, filled with *pneuma psychikon*, each charged with a specific function. The first ventricle, or cell, is devoted to the *sensus communis*. The sensory nerves run to it bringing tidings of the external world. The second ventricle is concerned with cogitation and the third with memory. There seems to be an ambiguity here that continues for well over a millennium into the times of the European Renaissance. It is the same ambiguity that Galen noted in his *De usu respirationis*. Is it the ventricle and its surrounding substance, or the pneuma within that forms the physical basis of the mental faculty? Is the *pneuma psychikon* or *spiritus animalis*, as it came to be known in Latin, the seat of the psychic function, or does this function reside in the substance of the brain leaving animal spirit the lesser task of merely in some way transmitting the latter's commands?

Nemesius, himself, recognizes this perplexity. '…when our foot hits a thorn', he writes,

> the hairs on our head immediately shiver and some have thought that the affection, or sensation of the affection, is sent upward to the brain. Yet if this account were true, it would not be the part that is cut that suffers pain, but the brain. It is therefore better to say that the nerve *is* the brain, for it is a part of the brain, which has psychic *pneuma* all throughout itself, just as iron that has been heated in the fire contains the fire; well, for this reason, then, wherever a sensitive nerve grows, this part has a share in sensation because of this, and it becomes sensitive. But perhaps there is nothing wrong about saying that what is sent upwards to the brain, the origin of the nerves, is not the affection but some kind of awareness and a report about the affections.[7]

Even at this early stage in the history of the 'hard problem' it is clear there is some dispute about the anatomical location of sensation. A variant of Nemesius's first proposal can be found as late as Erasmus Darwin, who promoted it in his widely-read text, the 1796 *Zoonomia*.[8] This thesis suggests, of course, that the *pneuma* is itself sensitive, not a mere messenger. Finally, Nemesius recognizes that the soul itself is incorporeal, arguing, rather as Aristotle argued for the immortality of the rational soul, that 'it is nourished by studies'.[9]

This perplexity about the location, because incorporeal, of the soul is also, famously, to be found in Nemesius's better known contemporary, St. Augustine of Hippo, who shows considerable familiarity with the so-called 'natural part' of

[6] Indeed, a slightly earlier but less well-known version of tripartite ventricular psychology was published by Posidonius of Byzantium in the middle of the fourth century.

[7] Nemesius, 8; in Sharples and van der Eijk, 2008, p. 122.

[8] Darwin 1801, vol. 1, p. 28: '… our ideas are animal motions of the organs of sense.' See, also, Smith 2005.

[9] Nemesius, 2; in Sharples and van der Eijk 2008, p. 55.

medicine (our physiology) which he probably picked up from the medical men of his time.[10] In some ways his understanding of physiology looks back through Galen to the earlier work of the great Alexandrians. It seems that he probably received this from his friend Helvius Vindicianus who was the author of several medical treatises.[11] Like Herophilus and Erasistratus, but unlike Galen, Augustine still believed that the arteries contained an air-like *pneuma*. In *De genesi ad litteram* he writes that 'air … spills out of the heart into the vessels we call arteries'.[12] It is interesting to note the sheer persistence of this ancient idea. We can still find echoes of it down into the seventeenth century of our era. Francis Bacon writes, for instance, that 'the lungs … through the *artire*, throat and mouth makyth the voice.'[13][my italics].

Erasistratean neurophysiology is also evident in other parts of Augustine's medical writings. The nerves, he says, form fine conduits between the body's surface and the brain. Thus in *De genesi ad litteram* he writes that 'Fine channels [*tenues fistulae*] run not only to the eyes, but also to the other senses, the ears, the nostrils, the palate, permitting the senses of smell, taste, touch etc.'[14] The eyes and the other sense organs are, he continues, the 'body's doors' [*fores corporis*] and are connected to the brain by hollow nerves which contain a material air-like spirit [*Deus hunc flatum fecerit, quae anima dicitur:* 'God has made this wind which is called the soul'].[15] This 'wind', he continues in the same section of *De genesi ad litteram* forms an intermediary between the 'control centres' in the cerebrum and the sense organs and/or members [*Et aer, qui nervis infusus est, paret voluntati, ut membra moveat, non autem ipse voluntas est*: 'and the air which is infused into the nerves obeys the will so that it moves the members without itself being the will']. Finally, Augustine comes to the brain itself. Like Nemesius (but, again, unlike Galen) he accepts a three-ventricle 'neuropsychology'. The sensory nerves run to an anterior ventricle (the seat of the *sensus communis*), the second ventricle ('towards the base of the brain') is concerned with movement, whilst the third ('between the two others') has to do with memory.[16] Clearly Augustine adopts a different and perhaps yet more fanciful schematic then the usual tripartite system adopted by the majority of the medievals (see below).

Augustine, as would be expected of a Christian theologian, is, however, more interested in man's immortal soul than in the physiological plumbing which allows it to interact with the body. Most importantly he argues that, unlike the *pneuma zotikon* and *pneuma psychikon*, the rational soul is immaterial. He adduces a number of reasons for this conclusion. Perhaps the most telling and the most important is his

---

[10] See Bardy 1953.

[11] An account of the debt Augustine owes to Vindicianus' lost medical treatises is given by Agäesse and Sogignac in St Augustine's *Oeuvres*, vol. 48, pp. 710–714.

[12] Augustine, *De genesi ad litteram,*7, 13, 20; in Agäesse and Solignac, 1972.

[13] Bacon 1620, p. 46, §199.

[14] Augustine, *ibid*., 7, 13, 20; in Agäesse and Solignac, 1972.

[15] Augustine, ibid., 7, 19, 25; in Agäesse and Solignac, 1972.

[16] Augustine, ibid., 7,18, 24; in Agäesse and Solignac, 1972.

concept of '*intentio*', the phenomenon of 'tension', 'attention', 'concentration'.[17] In another important work, *De quantitate animae*, he developed a second argument for the incorporeality of the rational soul, which was a powerful critique of the meaning of 'size' or 'quantity'. He points out that we use the same term to describe something fundamental in both physical and moral discourse. But we mean something different. When we describe Hercules, he writes, as having a great physique and when we describe him as having great principles we are alluding to categorically different concepts. It would make no sense to attempt to weigh or use a ruler to measure the latter. Similarly our imaginations can contain great cities, or tiny ants: physical size is of no significance. '*Magna quaedam, crede mihi, magna, sed sine ulla mole de animo cogitanda sunt*' [The soul must be pictured as something great, believe me, great, but without physical dimensions].[18]

Finally, in a closing argument to show that the 'soul' is incorporeal, Augustine considers how it is that a cast lizard's tail remains active and may even distract a predator, and how it is that a centipede or 'worm' may be cut in two and the divided parts still crawl away and these parts may, in turn, be divided several times more. How could this happen if the soul were extended through the body? Was the soul divisible? Or were there unnumbered separate souls located in each part? Augustine's solution refers back to the Stoic notion of 'tension' mentioned above. He resorted to an explanation which might have pleased Aristotle. This points out that as a word, say 'Mediterranean', may be abbreviated as 'med' without losing its meaning, so may a worm's body be divided without its soul being divided. Thus in *De quantitate animae* he writes that:

> ……… just as the worm as a whole occupied more space than any part of it, so a greater span of time is taken up in saying '*Lucifer*' than if one were to say only '*Luci*.' Hence, if this latter 'lives' in virtue of its meaning, in the diminution of time brought about by the division of that sound, while the meaning itself is not divided—for not the meaning, but the sound, was extended in time—then we should judge in the same way of the worm with its body cut to pieces: that, although a part, by the simple fact that it is a part, lives in a smaller space, still the soul is not at all divided, nor has it been reduced in a reduced space, notwithstanding that it simultaneously dominated all the members of the whole living body, when they were extended over a larger space. The soul, you see, occupied not space, but the body which it controlled.[19]

Augustine died at the age of 76 in 430 AD. His life encompassed the final break-up of the Western Empire and, in particular, the sack of Rome by Alaric and the Goths in 410 AD. The cutting edge of biomedical learning and research then moved eastwards to the Arabic lands of the Islamic ascendency. Here the Islamic physicians used translations of the master-works of classical antiquity, especially Aristotle and Galen, as bases for their understanding of the physiology of the human body.

[17] Letter to Jerome (transl, O'Daley 1987), chapter 4: '[the soul] is spread throughout the entire body which it animates, not through any local extension, but as a kind of vital tension.' This idea is common in Stoic philosophy.

[18] Augustine, *De quantitate animae*, XIV, 24.

[19] Ibid., 32, §68.

They mostly accepted the three-cell psychophysiology of late-classical times. Haly Abbas, for instance, in the second half of the tenth century C.E. and one of the most influential of the great series of Islamic physicians, accepted that animal spirit is confined within three cerebral ‘cells’. The first of these cells was concerned with the *sensus communis* and imagination, the second with cogitation and the last (most posterior) with memory. Once again the passage from the middle to the posterior ventricle is controlled by a worm-like structure or vermis, which acts as a valve.

An interesting interpretation of ventricular psychology can be found in another tenth-century publication, the influential treatise entitled ‘the difference between the spirit and the soul’ published by the prolific Arabic writer, Qusta ibn Lūqā (864–923). He uses ventricular psychology and Galen’s ‘worm’ to account for what he alleges to be a common behaviour: ‘Those who want to remember’ he writes, ‘look upwards as this raises the valve between the posterior and middle ventricles and spirit can get through the passageway and retrieve memories from the posterior ventricle; on the other hand those who want to think look down as this closes the valve and protects the cogitative faculty in the middle ventricle from unwanted memories’. Qusta ibn Lūqā’s treatise, translated into Latin as *De differentia spiritus et animae* was influential in the Latin West into the thirteenth century.[20] His theory implying that animal spirit is subject to gravity and thus like other objects has ‘weight’ seems to indicate that the physicality of animal spirit was not in doubt during the medieval period.

When we come to the so-called ‘12th century Renaissance’, we find that the first European medical text that has come down to us, that written by Nicolai the Physician, carries on the ventricular tradition. The brain, Nicolai writes, ‘is divided into three cells, the *cellula phantastica* in the anterior part of the head, the *cellula logistica* in the middle, the *cellula memorialis* in the posterior part.’[21] Interestingly, from the point of view of this chapter, Nicolai is concerned about the relative proportion of spirit in the ventricles to surrounding marrow (or white matter). He writes that the role of spirit is ‘to provide sensations and motion in the member’, whereas that of the marrow is ‘to permit free perception of diverse forms and shapes’. We are also told that the nerves carry animal spirit to all the members, endowing them with sensation, motion etc. The sensory nerves run to the first ventricle, the motor nerves depart from the third. Here, again, we have our perplexity. Are the spirits sensitive or do they merely transmit information to and from the substance of the brain? Or both? There is also a contrast: the brain’s marrow is perceptive while the psychic pneuma in the members confers sensitivity.

As the medieval period lengthened, with its dearth of practical dissection, ventricular psychology gradually swung free from whatever anchorage in anatomical reality it

[20] The Latin translation of Qusta ibn Lūqā’s treatise was one of the set books in Natural Philosophy in Paris in 1234. A more comprehensive account of ibn Lūqā’s theory may be found in Lokhorst and Kaitaro, 2001.

[21] *Anatomia Magistri Nicolai Physici*.

had had in earlier times. In an exhaustive paper Manzoni lists over 60 different medieval accounts of cell theory.[22] Many of these had no contact with anatomical reality and most seem to have been derived from Latin translations of Avicenna's very influential works. As Fig. 1.1 shows, the original tripartite schematic favoured by Nemesius and Augustine has been developed by subdividing the first two 'cells'.

Avicenna's elaboration of the original tripartite schematic is similar to diagrams in Roger Bacon's *De scientia perspectiva* (copied in 1428) and in Albertus Magnus's *Parvulus philosophie naturalis* (copied in 1473).[23] The latter was one of the most influential neuropsychologies of the Latin West in the mid-thirteenth century. Modifying Avicenna, Albertus divided each of the three classical ventricles into two, allowing his schematic to house six mental faculties: *sensus communis, imaginatio/estimatio, fantasia,*[24] *cogitatio, reminiscentia, memoria* (Fig. 1.2).

The psychophysical ambiguities are continued into Albertus' work. The spirit contained in the ventricles is described as 'vaporous and luminous' and, after a heavy meal, vapours ascend to the brain so that 'they block the paths of the animal spirits that minister to sense and motion and prevent the animal power from reaching the exterior senses'[25] and sleep ensues: very materialistic! The first faculty of the mind is, he says, located in the wall of the first ventricle. It is able to integrate the deliverances of the various senses and synthesize a unified object—its colour, texture, shape etc. How this occurs, now known as the 'binding problem', is as ill-understood today as it was a millennium and half ago. But note: Albertus believes that it takes place in the substance of the brain, not in the ventricle itself. Behind this faculty is '*imaginatio/estimatio*'. This is the power to retrieve images of objects after they have been withdrawn (= retentive imagination).

Nevertheless, the spirit carries the outcome of this computation onwards into the middle ventricle. How it becomes impressed with the image computed in the walls of the first ventricle is not explained—and remains another perplexity. It may be that spirit is impressed with the image much like wax with a seal. This would be consistent with Albertus' treatment of the animal spirit in the sensory nerves. This spirit, Albertus explains, flows out from the brain's ventricles through the nerves to the sense organs. Here it is impressed with sensory stimuli, much like wax can be imprinted with a seal, and these "*species sensibiles*" are carried back to the first of

---

[22] Manzoni 1998

[23] Clarke and Dewhurst 1972

[24] Bacon: *Opus Majus*: part V, chapter 2 (Bacon and Bridges 1897): 'in the anterior part of the first cell is the sensus communis. This takes cognisance of, and distinguishes, the impression brought by each special sense. But it is unable to retain these impressions, being loose and slippery. In the back part of the same cell there is therefore the organ of imagination, which, being neither too moist or too dry, can retain and store up the material received by the *sensus communis*. Avicenna [he writes] cites, as an example, a seal, the image of which water readily receives but does not retain owing to its superabundant moisture; wax, however, retains the image very well owing to its tempered moistness with dryness. Wherefore, he says, it is one thing to receive, another to retain, as is clear from these examples.…

[25] Albertus Magnus: *Questions. IV, 9*. (in Resnick and Kiitchell 2008, p. 163)

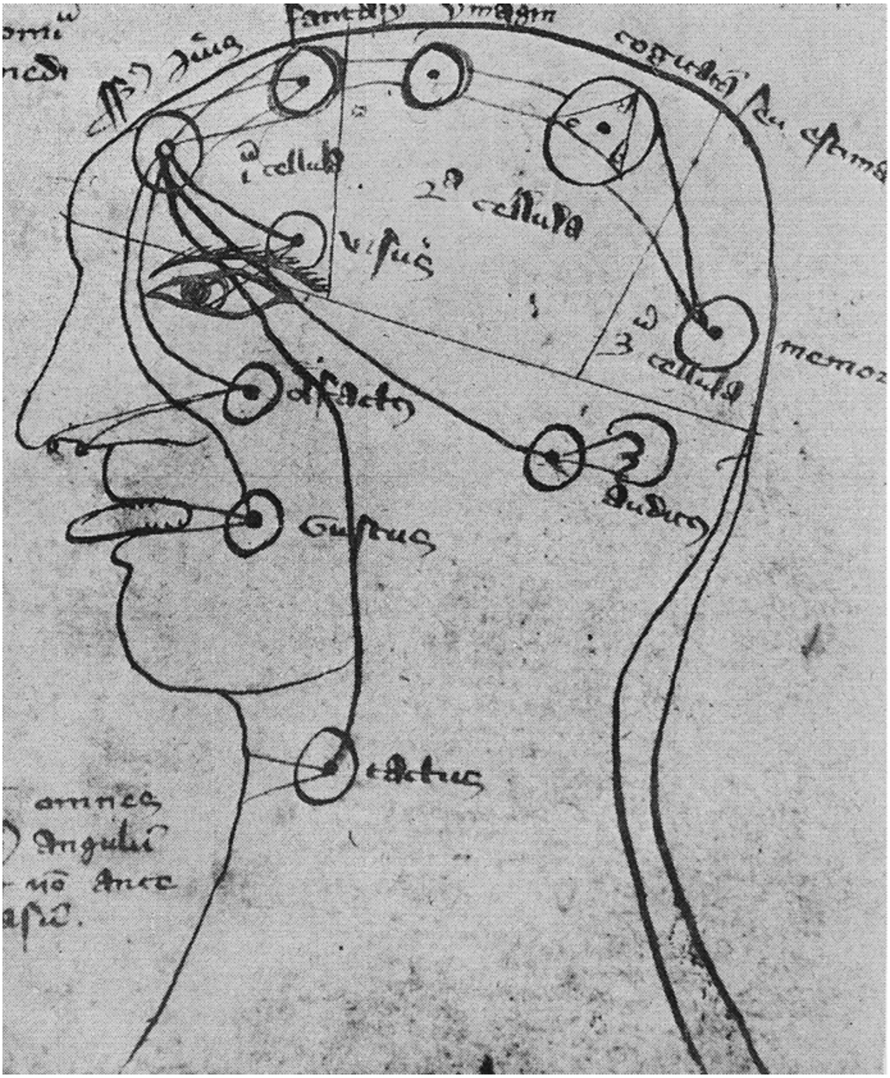


Fig. 1.1 Diagram at the end of Avicenna's *De Generatione Embyonis* (probably copied in 1347). The figure is merely captioned as 'this is the anatomy of the head for physicians'. It shows five interconnected 'cells'. The anterior cell is connected to the various senses: 'tactus', 'gustus', 'olfactus', 'auditus', 'visus'. It houses the 'sensus communis'. Successively, behind the sensus communis, are cells labelled 'fantasia', 'imaginativa', 'cogitativa seu estimativa' and 'memorativa'. The old tripartite schematic is also shown with vertical lines, perhaps added by the copyist. Although the vermiform valve is not shown the distinction between the functions of 'imaginativa' and 'cogitativa seu estimativa' is indicated by the triangle included in the latter 'cell' (From Clarke and Dewhurst 1972)

Fig. 1.2 Albertus Magnus' very well-known figure in his *Philosophia naturalis* shows a divided ventricle schematic. 'Sensus communis' and 'imagination/estimation' are located in the anterior ventricle; 'fantasia' and 'cogitatio', in the middle and 'reminiscentia' and 'memoria' in the posterior. See text

the brain's cells.[26] How does this compare with modern notions of information and information transfer? Information about touch, for instance, is carried by the appropriate nerve fibres in the form of a tattoo of impulses to a particular region of the somatosensory cortex and there the mysterious transmutation into a feeling occurs. Information about vision, according to David Marr, flows from primal sketch, through 2.5D sketch to 3D model.[27] But how this happens in the brain we do not

[26] Is shape a material concept? It cannot be atomised! The idea that the sensory nerves carry images of the world outside to the brain is, in fact, very ancient and can be found in, for example, Hunayn's *Art of Medicine* (c.850 AD) where he combines it with another Galenic idea: that sensory nerves are 'soft' whilst motor nerves are 'hard'. Whereas the sensory nerves are flexible and thus can accommodate the imprints of sensory stimuli, the 'hardness' of the motor nerve allows a percussive force to be delivered via the contained animal spirit to the muscles (for further information see Smith et al. 2012).

[27] See Marr 1976.

know. The onward flows to recognition and recollection are even more mysterious to contemporary neuroscience than they were to the great Albert.

The second cell in Albert's schematic is devoted to *fantasia* and *cogitatio*. The faculty of *fantasia* allows humans to envision mythological or unmet entities, such as unicorns, satyrs, basilisks, centaurs, sirens, and the like. *Cogitatio* is closely related to *fantasia* but involves the application of rational thought. Thus, the faculty of *fantasia* may toy with the idea of the three angles of a plane triangle adding up to more or less than 180°; *cogitatio*, on the contrary, determines that the angles of a plane triangle cannot add to anything other than 180°, and develops from this premise the other axioms of Euclidean geometry.

Albertus, like Aristotle and his Arabic predecessors, sees this transference from fantasy to rational thought as passing across the barrier from the material to the immaterial world, from the world of particulars to that of universals. This transference from the world of corporeal things to that of the intellect is controlled, according to Albertus, by Galen's worm-like valve. Here, however, the valve has been shifted forward to the centre of the middle ventricle. It is also interesting to note that in the medieval period '*pinea*' (or pine cone) began to be applied to this 'valve' and a number medical treatises asserted (Galen must have turned in his grave) that it was the pineal grand that was to be identified with this valve.[28] However, others, like Mondino dei Luzzi in the fourteenth century, took the valve to be the 'choroid plexus' in the fourth ventricle. Mondino describes it as a 'blood-red substance similar to a long worm', a description that is consistent with the structure we now know as the *tela choroidea*.

Finally, the third ventricle (our fourth) is associated with the faculties of *memoria* and *reminiscentia. Memoria* stores the outcome of the progressive work of the two anterior cells, while *reminiscentia*, an actively directed faculty, is that which recalls memories. It is important to note that these six faculties are regarded as located in the solid walls of the cells, while transference of "information" from one cell to another is accomplished by the *spiritus animalis*. The intellectual activity itself occurs, still mysteriously, in the seemingly homogeneous substance of the brain.

This separation between mental activity and the transmission of the results of that activity is continued into the most widely-read compendium of contemporary knowledge, in the late Middle Ages: Gregor Reisch's *Margarita philosophica* or *Pearl of Wisdom*. This was published in the late fifteenth century and re-published many times in the following century.[29] His well-known figure (Fig. 1.3) shows that the various mental faculties are labelled against the ventricular walls.[30] But note that the vermis has lost its anatomical position and now lies between the first and middle ventricle! This position is also shown in Robert Fludd's famous illustration in his

[28] In 1484 the Florentine poet Luigi Pulci was denied Christian burial for declaring that the soul was 'no more than a pine nut in hot white bread' (Brown 2010, p. 11).

[29] Variants of Reisch's figure appear in many sixteenth century texts and were even used in some phrenological works as late as the early nineteenth century (Clarke and Dewhurst, p. 39).

[30] It may be, of course, that this is no more than a typographical convenience. It is, nevertheless, to found in all the later sixteenth-century diagrams.

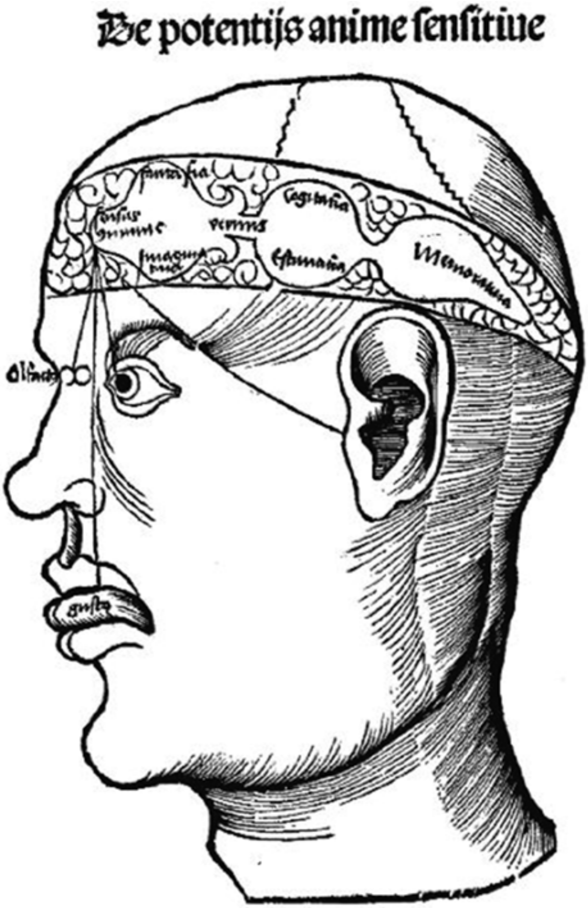


Fig. 1.3 Gregor Reisch's figure in the 1503 *Margarita philosophica*. The lines from the tongue, nose, ear and eye converge, as in Fig. 1.1, on the sensus communis in the anterior ventricle. Reisch also inserted fantasia and imagination into the walls of this ventricle. Note that the vermiform valve now connects the first ventricle to the second. The latter ventricle is labelled on its walls as concerned with cogitation and estimation (or judgement). The third and posterior-most ventricle is concerned with memory. Note the diagrammatic representation of the cerebral substance around the ventricles and the coronal suture lined up between the anterior and middle ventricles

1619 *Microcosmi historia* where it is represented by a sinuous line connecting those two ventricles.[31] Psychology and, in Fludd's case Rosicrucian metaphysics, has come to dominate anatomy!

At about the same time that Reisch published his compendium, Leonardo da Vinci was also investigating and illustrating human anatomy. Although his earlier figures showed the conventional three-ventricle schematic (Fig. 1.4a), his last

[31] For information about Robert Fludd and the seventeenth century Rosicrucians see Yates 1972.

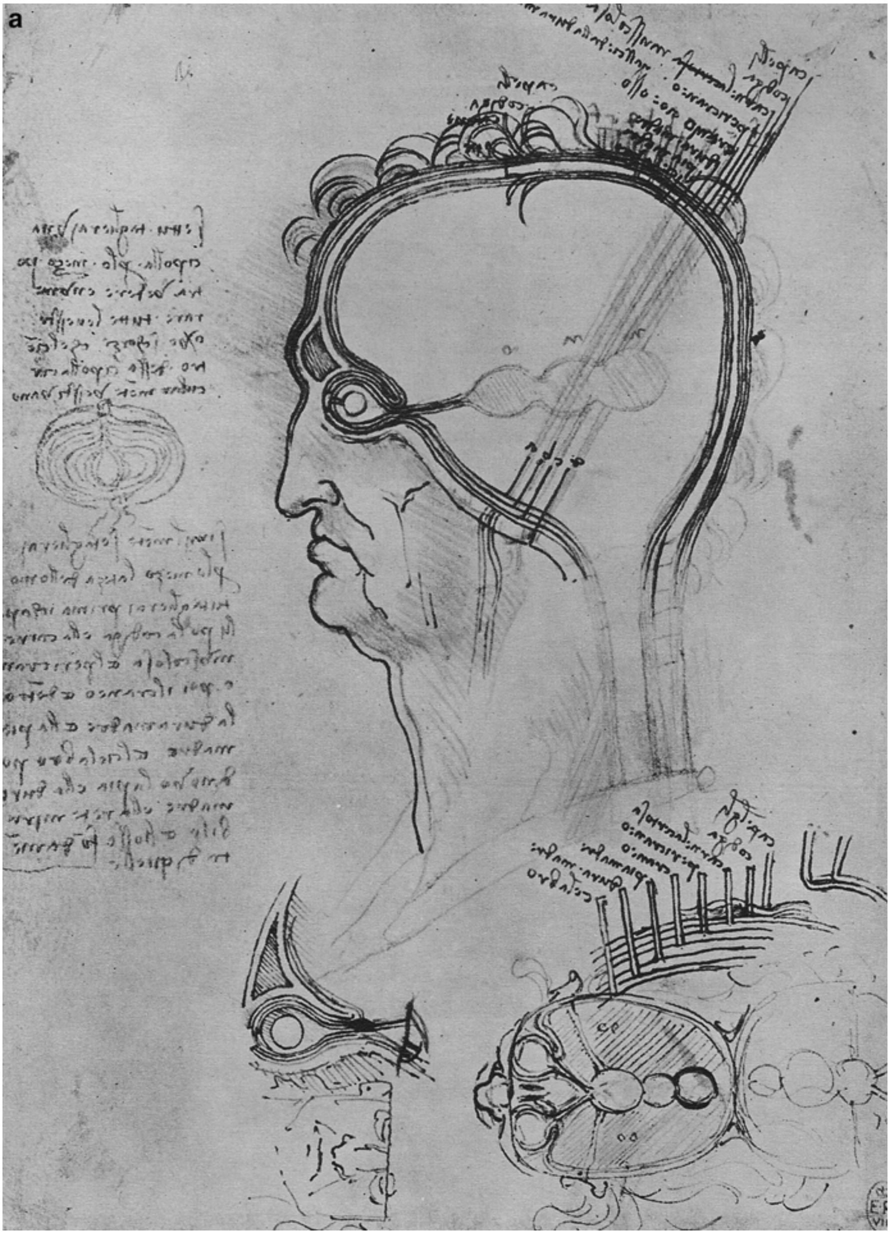


Fig. 1.4 (**a**) Leonardo da Vinci (c. 1493–1494). The ventricles follow the ancient tripartite schematic. On the *right side* of the bottom of the picture Leonardo has represented a horizontal section of the brain showing again the three (imaginary) ventricles. The top part of the skull is hanged backward. (**b**) Leonardo (c.1508–1509). The figure shows the wax-embedded ventricles in sagittal and horizontal section. The notes give detailed instructions of the injection technique and this is also illustrated in the small drawing in the *lower right* hand corner of the sheet

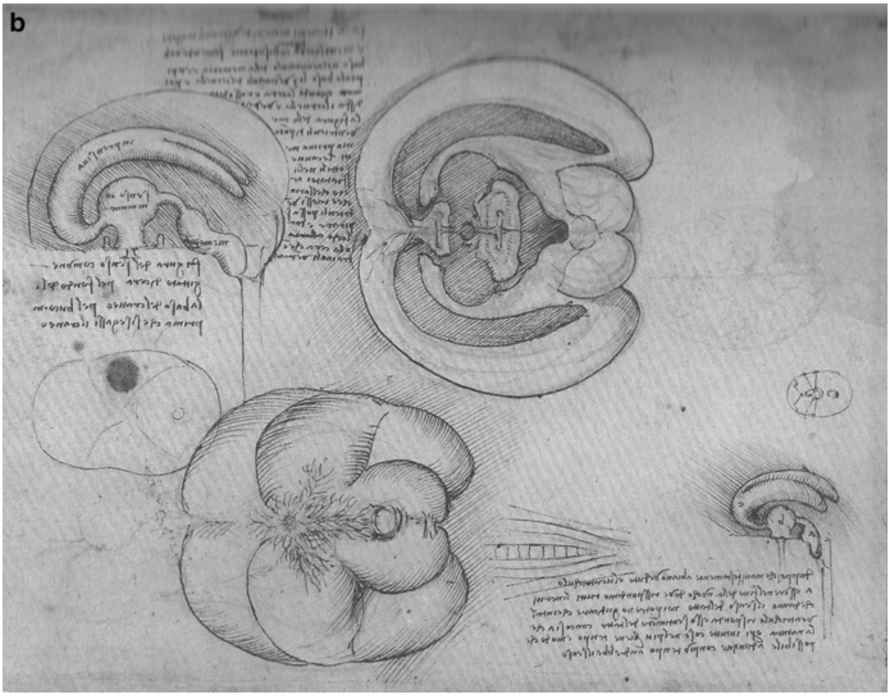


Fig. 1.4 (continued)

figures (probably drawn from a combination of wax-embedded ox and human ventricles[32]) were far more realistic than anything since Galen (Fig. 1.4b). It is interesting to note, moreover, that he labels the ventricles in an unorthodox manner: the first he labels *imprensiva* (perception), the middle *sensus communis* and third *memoria*. The central position of the *sensus communis* is likely to be because of Leonardo's perception of its centrality and also to be where he believed the majority of the sensory nerves terminated. In spite of his original and questing intellect, he still retained the medieval mixture of the psychological and the material. He writes, for instance, that 'mental matters which have not passed though the *sensus communis* are vain, they give birth to no other truth than what is harmful. And', he goes on, 'because such discoveries spring from a poverty of intellect, their authors are always poor and, if they were born rich, they will die poor in their old age'.[33]

With Vesalius in the next century we are aware of the momentous changes sweeping through the worlds of science (natural philosophy) and philosophy in Europe. In

[32] See Clayton 1992.

[33] O'Malley and Saunders 1983, p. 338. Leonardo goes on to instance the attempts of Alchemists to turn lead into gold and the endeavours of Necromancers to communicate with the dead as leading only to the poor-house.

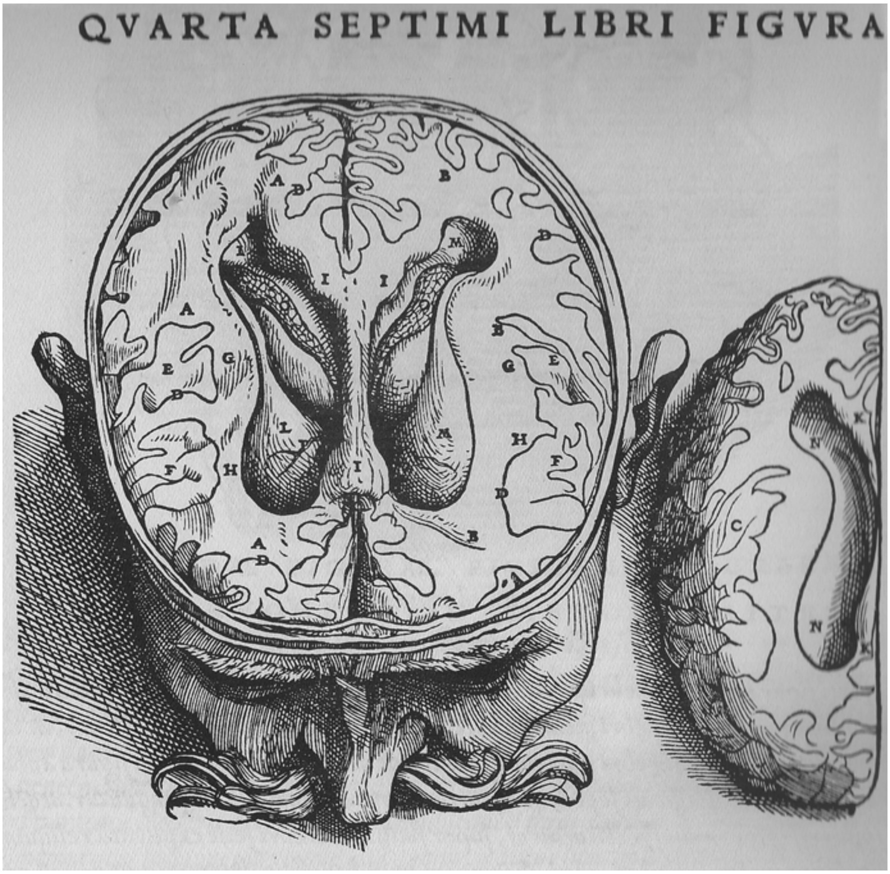


Fig. 1.5 Andreas Vesalius (1543): *Fabrica*. Horizontal section through the brain to show ventricles

the 1543 *Fabrica* at least, he no longer sees the ventricles as imaginary cavities designed to provide a physical basis for faculty psychology. His careful anatomy shows them very much as we now know them to be (Fig. 1.5). He recognized that they are filled with a watery fluid. Furthermore he finds no crucial difference (except in size) between the brains and ventricles of humans and those of infra-human mammals and, indeed, birds. What is he to make of this? Perhaps unnerved by observing so many anatomical similarities between thinking and non-thinking organisms, he merely writes, that "whatever likelihood should arise in my mind could not be set down without damaging our most holy faith."

I cannot end this all-too-short survey without reference to René Descartes who, in the next century, in some ways marks an end-point to this whole millennium-long story. Although not a trained anatomist, neither was he a mere armchair theoretician. In a letter to Mersenne in 1639 he writes that while composing *L'Homme* he

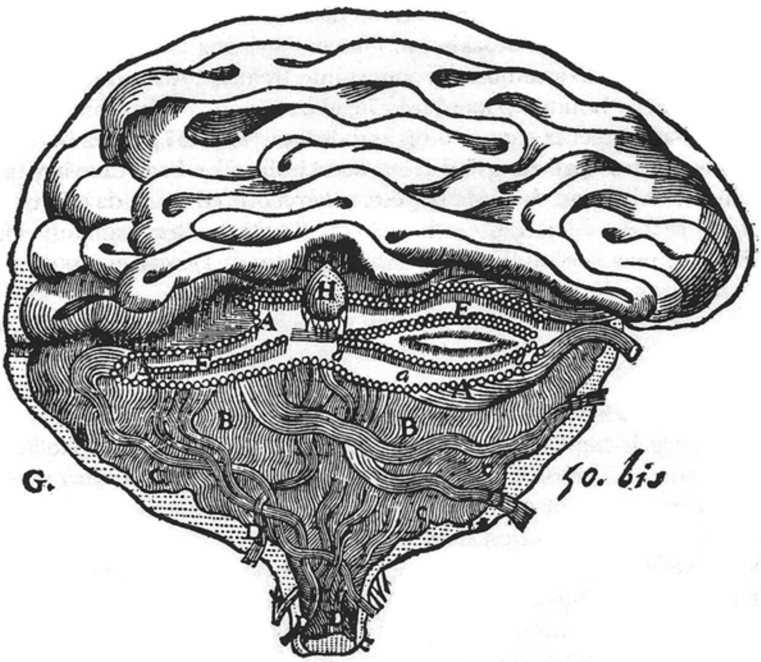


Fig. 1.6 René Descartes (1632/1662): *L'Homme*. Gland H floats in the cerebral ventricle buoyed up in animal spirits jetting from its lower surface. This figure was drawn by von Gutschoven after Descartes's death in 1650

consulted both Galen and Vesalius.[34] It is also known that in the 1620s he was close friends with Vospicus Plempius, a Galenist physician, who went on to become Professor of Medicine and then Rector of the University of Louvain and who published an influential text, *Fundamenta Medicinae*, in 1638.[35] In addition to Galen and Vesalius there is evidence that Descartes was familiar with Caspar Bauhin's *Theatrum anatomicum* (1605/1621).[36] He was also sufficiently interested, when in Amsterdam, to frequent the butchers' shops to carry off sheep's heads to dissect their brains at leisure in his lodgings. Nevertheless, the hydraulic neurophysiology which he was ultimately to publish in *L'Homme* looks back to the diagram-makers of medieval times. Although there is no trace of the traditional three ventricles, Gland H swings freely in a central cerebral chamber filled with animal spirits (Fig. 1.6).

[34] Letter, 1639.

[35] In the first (1638) edition Plempius strongly disputed Harvey's theory of the circulation of the blood. By the time of the second (1649) and third (1654) editions he had, however, revised his opinion and become a strong supporter of the Harveyan theory. By this time he had fallen out with Descartes and included an appendix in which three theologians asserted that Descartes' philosophy was incompatible with the Faith and that his system of medicine was dangerous to health.

[36] Britol-Heperides 1990.

Is this a genuflection to Galen's 'red worm'?[37] It receives on its surface impressions from the sensory nerves, most importantly the optic fibres from the retina, and the conscious will is able to move it to and fro so that animal spirits are directed down appropriate motor nerves to actuate the appropriate behavioral muscles.

The animal spirits in the ventricles and nerves are seen as mere messengers and above all physical. Their age-old psychophysical ambiguity is at last resolved. Coleridge was right when he accused Descartes of being the first to make all nature lifeless.[38] Mentality is somehow confined to the little gland, both for both perception and willing. But how does an immaterial, non-extended substance, *res cogitans*, affect the material, extended substance, the '*res extensa*' of the gland? Descartes does not say.[39]

Descartes' analysis is in some ways similar to that of St. Augustine a millennium before.[40] This similarity was noticed soon after Descartes published the *Discourse on Method* where his central insight '*cogito ergo sum*' seems remarkably similar to Augustine's '*si enim fallor, sum*' (for if I err, I am).[41] Andreas Colvius (or Colville), a minister in Dordrecht, drew this to Descartes' attention and René went immediately to the town library to check the reference.[42] He wrote thanking Colville for pointing out the similarity. But the 'hard problem' of the relation of mind, *res cogitans*, to matter, *res extensa*, is far more difficult than in Augustine's day. For by Descartes' time, in the early seventeenth century, the Galilean revolution in the physical sciences was in full spate. Johannes Kepler considered the world to be a giant clockwork and Descartes agreed, writing that 'l'univers entier est une machine où tout se fait par figure et mouvement'[43] and that all the functions of the human body 'follow from the mere arrangement of the machine's organs every bit as naturally as the movements of a clock follow from the arrangement of its counterweights and wheels.'[44] The concept of the inanimate was far from being so sharply defined in the fifth century. The problem of how the 'mind' interacted with the 'brain' was accordingly not so pressing for St Augustine.

Perhaps we, who still live in the long aftermath of the scientific revolution of which Descartes was so much a part, should copy his (Descartes') example. Just as he sought to find a place for mind in the Galilean world picture, so we should seek

---

[37] Lokhorst and Kaitaro (2001) who have made a careful study of Descartes' sources conclude that his pineal theory is largely original with him.

[38] Coleridge, *Philosophical Lectures*, 1818–1819: 'Descartes was the man who made Nature utterly lifeless... and considered it as a subject for purely mechanical laws'. See Coburn 1949, pp. 376–378.

[39] In a letter to his much admired Princess Elizabeth of Bohemia he writes that it is a waste of life to spend too much time over the problem (Descartes, June, 1643: in Anscombe and Geach 1954, p. 282).

[40] See Smith 1998.

[41] Augustine, *De civitate Dei*, 10, 26.

[42] Kenny 1970, p. 83.

[43] Quoted Crombie, 1959, vol. 2, p. 86.

[44] Descartes 1633/1662, trs Hall, p. 202.

to find a place for mind in the world of the new physics initiated by Bohr, Schroedinger, Heisenberg and others in the twentieth century. Twentieth and twenty-first century physics has redefined once more, and in a radical fashion, our concept of physical reality.[45]

1.1 Concluding Remarks

In this chapter we have seen how the hard problem, as presently understood, slowly emerged from relative obscurity, some would say non-existence, in ancient thought. We have seen how, from the first, there was a distinction between 'sensitive' and 'rational' 'souls'. We have seen how from Galen onwards there was confusion over whether the 'animal spirit' in the ventricles of the brain and in the hollows of the nerves was merely a material messenger or was itself sensitive, or both. We have seen how the major mental faculties were conceived to be located (insofar as they possessed spatial location) in the substance, the 'marrow', of the brain, and communicated with animal spirit through the walls of the ventricles. We have seen how, with the scientific renaissance of the sixteenth and seventeenth centuries, the true nature of the brain's ventricles became apparent and the medieval faculty psychology with which they had become confused was, perforce, discarded. At the end of this revolutionary period René Descartes sought to clarify the neuropsychological dilemma—between what was the province of physical science and what was the province of the mind—in the form of his famous dichotomy. This, of course, is not the end of the story. Faculty psychology lived on in Robert Fludd's Rosicrucian fantasies and in the more sober imaginings of nineteenth century phrenologists. However, the next chapters in this book will follow out the repercussions of Descartes' radical idea in the history of neuroscience.

References

Albert the Great (2008) Questions concerning Aristotle's "On Animals". In: Resnick IM, Kitchell KF (trans) The Fathers of the Church, mediaeval continuation. Catholic University of America Press, Washington, DC

Anscombe E, Geach PT (1954) Descartes: philosophical writings. Nelson, London

Bacon F (1620) Sylva Sylvarum, unpublished mss

Bacon R, Bridges JH (1897) The Opus Majus of Roger Bacon, 2 vols. Clarendon Press, Oxford

Bardy G (1953) St Augustine et les médecins. L'Année Théologique Augustinienne 13:327–346

Bauhin C (1605, 1st edn, 1621, 2nd edn) Theatrum anatomicum. Mattaei Bekeripp, Frankfurt

Britol-Heperides A (1990) Le Principe de vie chez Descartes. Vrin, Paris

Brown A (2010) The return of Lucretius to Renaissance Florence. Harvard University Press, Cambridge, MA

[45] See Chaps. 16, 17 and 18, this volume.

Clarke E, Dewhurst K (1972) An illustrated history of brain function. Sandford Publications, Oxford
Clayton M (1992) Leonardo da Vinci: the anatomy of man. Museum of Fine Arts and Bullfinch Press, Houston
Coburn K (1949) Philosophical lectures of Samuel Taylor Coleridge. Pilot Press, London
Darwin E (1801) Zoonomia; or the laws or organic life, 3rd edn. J. Johnson, London
Descartes (1633/1662) De homine. Leiden; Hall TS (trans) (1972) Treatise of man. Harvard University Press, Cambridge, MA
Furley DJ, Wilkie JS (1984) Galen on Respiration and the Arteries: an Edition with English Translation and Commentary of De usu respirationis, An in arteriis natura sanguis contineatur. De usu pulsuum, and De causis respirationis. Princeton University Press, Princeton
Galen. On the natural faculties; trans. Brock AJ (1916) Heinemann, London
Galen. De usu partium corporis humani [On the usefulness of the parts of the body]; trans. May MT (1968) Cornell University Press, New York
Kenny A (1970) Descartes: philosophical letters. Clarendon, Oxford
Lokhorst G-J, Kaitaro TT (2001) The originality of Descartes' theory about the Pineal Gland. J Hist Neurosci 10:6–18
Manzoni T (1998) The cerebral ventricles, the animal spirits and the dawn of brain localisation of function. Arch Ital Biol 136:103–152
Marr D (1976) Early processing of visual information. Philos Trans R Soc Lond B Biol Sci 275:483–524
O'Daley G (1987) Augustine's philosophy of mind. Duckworth, London
O'Malley CD, Saunders JB (1983) Leonardo on the human body. Dover Publications, New York
Siegel RE (1968) Galen's system of physiology and medicine. Karger, Basle
Sharples RW, Van der Eijk P (2008) Nemesius: on the nature of man. Liverpool University Press, Liverpool
Smith CUM (1998) Descartes' visit to the town library, or how Augustinian is Descartes' neurophysiology? J Hist Neurosci 7:93–100
Smith CUM (2005) All from fibres: Erasmus Darwin's evolutionary psychobiology. In: Smith CUM, Arnott R (eds) The genius of Erasmus Darwin. Ashgate, Aldershot, pp 133–143
Smith CUM, Frixione E, Finger S, Clower W (2012) The animal spirit doctrine and the origins of neurophysiology. Oxford University Press, New York
St. Augustine. De quantitate animae. In: de Labriolle P (ed and trans) (1948) Oeuvres de Saint Augustine, vol 5, Études Augustininennes. Desclée de Brouwer, Paris
St Augustine. De genesi ad litteram. In: Agäesse P, Solignac A (ed and trans) (1972) Oeuvres de Saint Augustine, vols. 48, 49, Études Augustinennes. Desclée de Brouwer, Paris
St Augustine. De civitate Dei; Healey J (trans) (1931) The city of God. Dent, London
Willis T (1681) The anatomy of the brain and nerves, facsimile edition of Cerebri Anatome. Englished by S. Pordage, edited by W. Feindel (1965) McGill University Press, Montreal
Yates F (1972) The Rosicrucian enlightenment. Routledge and Kegan Paul, London

Chapter 2
Return of the Repressed: Spinozan Ideas in the History of the Mind and Brain Sciences

William Meehan

We know ourselves to be part of the totality of nature.

—Spinoza, *Ethics* 4apdx32

2.1 Introduction

Our contemporary understanding of the "hard problem" in the mind and brain sciences is rooted in a general conception of science as the reductive study of physical matter that originates in seventeenth-century efforts to develop a viable alternative to the scholastic synthesis that had permeated western education and thought for the previous 500 years. Continental philosophers like René Descartes (1596–1650), as well as British Atomists and Empiricists like Thomas Hobbes (1588–1679) and Thomas Willis (1621–1675), sought to establish a new paradigm for natural philosophy, which included an understanding of the relationship between our psychological experience and a mechanist physics. For all of them, scientific accounts had to be limited to a kind of causal analysis in the modern sense that equates *cause* with what the scholastics would have called *efficient cause*, and they all rejected any appeal to final or teleological causality. These, of course, are all important values in contemporary science, and such seventeenth-century natural philosophers are easily construed as standing at the beginning of a relatively linear narrative connecting their insights to the contemporary paradigm. It is because of their inaugural role in the creation of contemporary science that they are remembered and that their ideas still seem reasonably familiar, in spite of the 300+ years that have elapsed since the beginning of the early modern period.

Benedictus de Spinoza (1632–1677) was engaged in the same early-modern quest for an alternative to scholasticism, but his approach was sufficiently different

W. Meehan (✉)
Independent Scholar, 10 Funston Avenue, The Presidio, San Francisco, CA 94129, USA
e-mail: Wmmeehan@sbcglobal.net

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_2,

from that of the Cartesians and Empiricists that an account of the role of his ideas in the history of science, and specifically of the mind and brain sciences, does not fit well into what might be considered the canonical narrative. As will be discussed below, his ideas were perceived by his contemporaries as far more radical than those of other natural philosophers, and as a result, they were condemned by both civil and religious authorities and marginalized within the intellectual community.[1] Recently, the limitations of the Cartesian/Empiricist approach have led many neuroscientists, particularly those working on consciousness and mind/brain relationships, to consider quite Spinoza-like models,[2] and one major researcher, Antonio Damasio[3] has explicitly acknowledged Spinoza as a major influence on his work. For the most part, however, the neuroscientific progress of the intervening centuries was accomplished without much reference to his work. The pattern of Spinozan influence in the history of the mind and brain sciences, thus, is less a linear narrative than what the Freudians might describe as a "return of the repressed." And, as the initial suppression of his ideas affected not only the mind and brain sciences, but also the development of the broader intellectual culture, Spinoza's philosophy is often considered more problematic for contemporary readers[4] than the more dominant ideas of thinkers like Descartes or John Locke (1632–1704).

Methodological reductionism and mind/body dualism, both of which Spinoza rejected, became central to classical scientific thinking, with mind/body dualism achieving such hegemony that alternatives to it—e.g., the idealism of George Berkeley (1685–1753) or the radical *materialism* of Julien Offray de La Mettrie (1709–1751)—were largely conceptualized in terms of a forced choice between incorporeal mind or inert matter. What Spinoza suggested as an alternative was a suite of ideas, the most important of which (for the present purposes) are holism, dual-aspect monism, and, derived from these two, a notion of mind/body relations that defines human affects as changes in the organism's ability to sustain itself. The intent of this chapter is the explication of these three ideas, as they appear both in Spinoza's philosophy and in subsequent neuroscientific work. To that end, the chapter is divided into two main sections: the first, an account of how they appear in Spinoza's thought, and the second, an effort to capture the historical return-of-the-repressed narrative that accounts for the reemergence of a Spinozan perspective in the mind and brain sciences in the latter part of the twentieth century.

2.2 Spinoza's Philosophy

When contemporary readers find Spinoza difficult to understand, it is largely because his ideas run counter to the hegemonic, Cartesian (and Empiricist) paradigm. The task of interpreting Spinoza is made somewhat easier, however, by the

[1] Israel 2001.

[2] Ravven 2003.

[3] Damasio 2003.

[4] Cf. Allison 1975.

fact that his major philosophic work, the *Ethics*, was a relatively explicit critique of Descartes: a circumstance that allows us to use our familiarity with the latter to elucidate the former by way of contrast. The most important point at which such a contrast can be made is at the level of first principles, which for Descartes begins with his famous: "I am thinking, therefore I exist": the "clear and distinct idea" at which he arrived by means of his equally famous reductive method of hyperbolic doubt.[5] This Cartesian *cogito* is a principle with a great many ramifications, but for our purposes, there are two that are of particular importance. First, it implies a distinction between the human mind, which it asserts to be knowable by direct intuition, and everything else that the mind encounters. And, second, it asserts that this self-evident mind is the mind of the individual knower. This, of course, is a remarkably solipsistic principle; Descartes can only resolve "the hard problem" of connecting this self-awareness to other forms of knowledge by way of separate and additional claims about the existence and goodness of God.[6]

2.2.1 Nature as a Whole

The Cartesian *cogito* was intended as an intuitively certain fact on which the rest of Descartes's philosophic and scientific work could be grounded. In today's intellectual climate, few sophisticated people expect the kind of certainty that Descartes sought, but this goal was a pervasive one among the early-modern thinkers who were looking to replace the once universally accepted scholastic synthesis. It is a goal unquestionably shared by Spinoza, who begins his *Ethics* with a series of what he considers intuitively certain definitions, the most important of which is that of Substance: something "whose concept does not require the concept of another thing from which it must be formed" (*E*1d3).[7]

The term *Substance*, like many of the terms used by Spinoza—as well as by Descartes and others of their contemporaries—is adapted from scholastic technical vocabulary and does not carry much meaning in contemporary discourse,

5 Descartes 1637/1996a.

6 Ibid.

7 de Spinoza 1677/1985, *E*1d3. Abbreviations for the subsections in Spinoza's *Ethics*: Definitions (d), Axioms (a), Postulates (pos.), Lemmas (l), and Propositions (p), which have Corollaries (c), Scholia (s), and Demonstrations (dem). In addition, there are introductions and appendices to some of the Parts. Citations in this paper will first list the Part number, followed by a letter designating what type of subpart is cited, followed by its number. Thus, *E*2p1 refers to the first proposition in Part II. Where more than one section is referenced, a comma should be read as "and." Thus, *E*2p1,2 refers to Propositions One and Two in Part II. Corollaries and Scholia are associated with propositions as in *E*2p32c1 (Part II, Proposition 32 Corollary 1) or *E*2p40s1 (for a Scholium). References to the Postulates, Axioms, and Lemmas that follow *E*2p13 will be indicated by an accent mark; thus *E*2a1' refers to the first axiom in this section. Subdivisions of Introductions and Appendices will be transparent on their face. Direct quotations in the text are from Curley's translation (de Spinoza 1677/1985). The translation in the epigram is my own.

particularly the discourse about consciousness. In this context, however, the precise meaning of the term is less important than the breadth of its connotations, the role it plays as the foundational element in Spinoza's philosophy, and the contrast between it and the minimalist "consciousness of self" upon which Descartes grounds his system; where Descartes begins with certainty of self, Spinoza starts with certainty about this all-encompassing Substance of which, he will later argue, all singular or particular things are parts or aspects. In his technical vocabulary, singular things—including conscious selves—are *modes* or *modifications* of the totality,[8] which is both logically and causally prior to any singular thing (*E*1p1).

The implications of these contrasting starting points are important. As has already been noted, the reductive *cogito* does not lead the doubter to knowledge of anything outside of itself, and as a result, Descartes is forced to introduce additional fundamental ideas including that of an omnipotent and benevolent God who serves as guarantor of the truth of all other ideas that can be described as clear and distinct.[9] Yet another fundamental idea—made necessary by the fact that the thinking self is arrived at only after doubting the physical self—is a mind/body dualism so radical that each part is conceived as separate and distinct substances. Spinoza's intuition of an all-inclusive Substance is one that precludes the possibility of there being anything outside of itself: including anything—like Descartes's God—that might have created this whole. (*E*1p6c). Substance, for Spinoza, is "the cause of itself" (*E*1d1).

In the scholastic terminology with which Spinoza's contemporaries would have been familiar, however, the condition of having the quality of being self-caused (*causa sui*) is a definition of God.[10] Thus, rather than being created by God, Spinoza says that the totality of which everything is a *mode* or part, *is* God; and, throughout the *Ethics*, he refers to this totality as "God or Nature" (*Deus sive Natura*). This equating of God and Nature, however coherent it may have seemed to the scholastically educated, is not unproblematic, and there has been considerable disagreement about how far Spinoza meant the idea to be taken. To his own generation and their successors for the next 100 years, his definition of substance as *causa sui* was not sufficient to outweigh his rejection of an independent Creator, and he was thus universally condemned as an atheist.[11] But, to the German Romantics who rehabilitated his reputation in the late eighteenth century, he is a pantheist and a mystic.[12] Present-day opinions on this question depend on variations in one's definition of pantheism, but I am inclined to accept the assessment of his contemporaries and read the theistic language as an artifact rooted in the technicalities of the scholastic terminology, which was the only philosophic language available at the time.[13]

Questions about Spinoza's atheism are, for the present purposes, arresting but not central. What is more important is the rejection of reductive method implicit in his

[8] de Spinoza 1677/1985, *E*1d5.

[9] Descartes 1637/1996a.

[10] See, e.g., Aquinas 1270; *Summa*, I, q.2, a.3.

[11] See Gullan-Whur 2000; Israel 2001; Nadler 2001.

[12] See DeCuzzani 1991; Goetschel 2004; Jacobi et al. 1916/2010.

[13] Cf. Nadler 2011.

notion of Substance. For Descartes and the vast majority of his contemporaries who adopted an explicitly atomist perspective,[14] natural philosophy was, at root, caused by the behavior of some kind of fundamental particles—atoms or corpuscles—out of which all natural things are made. All natural phenomena are reducible to such parts, and, ultimately, only intelligible in terms of them. Causality, for the majority of the early moderns, operated—as it were, from the bottom, up—with complex phenomena explained as the effects of simpler and more fundamental particles and processes, which were seen to be their causes. For Spinoza, however, because he begins with the self-evidence of substance, causality is as much a top-down as a bottom-up phenomenon, for natural objects are unintelligible except in terms of the whole. Because he, like all his non-Scholastic contemporaries, rejected the idea of teleological causes in favor of a mechanistic explanation for the interaction between singular natural objects, he does have a conception of the sort of linear causality that is more familiar to contemporary sensibilities. But the relationship of particular things to the totality of nature is, for him, more important than their interrelationships, and he only begins his discussion of singular or particular things in Parts II-V of the *Ethics*: after he has established his conception of the, causally active, totality of nature (*E*1p23,25), as clearly and completely as he feels he can.

The essential elements of his concept of Substance, in addition to the idea of its being *causa sui*, are that it exists necessarily (*E*1p7,11), is infinite and eternal (*E*1p8), as well as unitary (*E*1p12) and unique (*E*1p14). Most important for present purposes, it is essential for an understanding of Spinoza's conception of a unified totality—Mind and Body—that we realize he does not see them as separate substances as they are in Descartes.[15] Neither, however, are they to be thought of as separate parts of the totality. They are, rather, what Spinoza calls *Attributes*: "what the intellect perceives of a substance, as constituting its essence" (*E*1d4). His doctrine of attributes is a complex one, but at its core is the assertion that each attribute constitutes the whole essence of substance, and the difference between them is a difference in conception: Neither one is the cause of the other (*E*3p2)—any more than one side of a coin is the cause of its obverse—and to speak of either is to speak of the whole.[16]

In contemporary discourse, we are accustomed to thinking of mind/body relationships in terms of particular minds, whether of humans, primates, or animals in general. Thus, Spinoza's meaning when he ascribes Mind to Nature as a whole is not immediately self-evident. Jonathan Bennett[17] has argued that Spinoza wasn't really interested in the details here, but merely reasoned that, since human beings

[14] Technically, Descartes was not an atomist because he rejected the notion of a void. However, his idea that matter is essentially corpuscular has the same reductive implications as atomism, properly so called.

[15] Descartes 1637/1996a, 1649/1996b.

[16] In an insight that anticipates contemporary *n*-dimensional physics, he asserts that substance, being infinite, consists in an infinite number of attributes (*E*1d6), though the only two attributes human beings are aware of are Mind, or Thought, and Body (*E*2p1, 2).

[17] Bennett 1984.

think, we have to ascribe Mind to the totality of Nature in order to hold that human beings are natural things—which is to say modes of God or Nature. This is, I believe, a helpful way to bracket the question of Nature's "psychology," but it is also useful to consider Spinoza's assertion, in *E*2p7, that "the order and connection of ideas is the same as the order and connection of thing*s*." In this connection, Edwin Curley[18] and Don Garrett[19] have suggested that we think of Mind as the set of propositions that can be made about Nature. Spinoza holds, as it were, that Nature "makes sense," and the sense that Nature makes is itself natural[20]; and it is this natural "sense" that he calls Mind (with a capital "M").

2.2.2 Singular Things

Spinoza's most basic ideas about the nature of singular things are laid out in the set of Axioms and Postulates that follows *E*2p13s. This is a series of statements on the nature of bodies in which he asserts that bodies are either in motion or at rest, that they differ, one from the other, only in the speed and direction of their movement, and that the speed and direction of a moving body is determined by—and can only be understood in terms of—the given body's interaction with other moving bodies. By the term *bodies*, in this initial context, he means simple uncompounded singular things roughly similar to the fundamental particles of reductionist mechanics: "corpuscles" in Descartes's fluid mechanics[21] or atoms in the works of Empiricists like Thomas Willis,[22] who adopted Gassendi's[23] revised Epicurean mechanism in preference to Descartes's hydraulic model. But, this similarity with the fundamental particles of reductionist mechanics is deceptive, because Spinoza's simple bodies differ from corpuscles or atoms in two important ways. The first of these is that, consistent with his notions of top-down causality, singular things are modes of the totality and are in no sense discrete fundamental particles (*E*1p1). Secondly, in Spinoza's philosophy, the description of a specific moving body is not exhausted by an enumeration of the external causes of its speed and direction, because these causes do not account for the body's own inertia or momentum—the fact that "each thing, in so far as it is in itself, strives to persist in its being"—what he calls the body's *conatus* (*E*3p6).

For Spinoza, a body's momentum or *conatus* is intelligible in terms of the body itself; it is, in effect, the actual essence of any singular thing (*E3p7*). This is a crucial point for two reasons. First, it is an assertion that bodies are not inert as they are for Descartes and the Empiricists. Second, Spinoza will argue that particular minds

[18] Curley 1969.
[19] Garrett 2002.
[20] Meehan 2009.
[21] E.g., Descartes 1662/1996c; Gaukroger 1995.
[22] Willis 1683/1971.
[23] Gassendi 1972.

(with a lowercase "m") are the ideas of particular bodies (*E*2p13,13s), and by saying that a simple body is intelligible in terms of its *conatus* or essence, he is identifying the idea or "mind" of that body.

The notion that there could be a "mind" of a simple moving body or particle, in and of itself, may seem incoherent, but it is important to bear in mind that, for Spinoza, nothing is ever just in and of itself; it is always part of something else. Ultimately, singular things are modes of the totality of Nature and must be, like Nature itself, intelligible in terms of both Body and Mind. In addition, simple particular things exist as parts of more complex singular things. Spinoza's notion of bodies, thus, is not limited to simple bodies, and his conception of *conatus* is not limited to simple inertial movement. He tells us that any grouping of simple bodies moving in concert and in such a way as to be the cause of a single effect is a body with its own *conatus* in its own right (*E*2a2'). In an observation that anticipates one of the central developments of twentieth-century neuroscience, he argues that the *conatus* of such a complex body is the internal dynamics whereby it maintains a pattern of relations among its parts. At the level of organisms, this amounts to homeostasis, and at the level of animals, it includes not only the internal homeostatic activity, but also the actions that the animal takes to provide itself with whatever it needs to maintain its life and reproduce. The relation between parts and wholes connects every singular thing to the totality of nature in a nested hierarchy of progressively more complex patterns. In contemporary vocabulary, we might say that organisms and animals exist as parts of ecologies, ecologies are organized into worlds, worlds into universes, and so on to the ordered totality of Nature. At each level of organization, there is, for Spinoza, an internal dynamic, a *conatus* or striving to persist in being.[24] Each of these, to the extent that they are spatially extended, are bodies in Spinoza's sense of the word. And each of these "bodies" is intelligible in terms of its *conatus*, which, to him, means that for each "body" there is an idea, or "mind."

The notion that the "minds" of such bodies and dynamic systems are the "ideas" of those phenomena is yet another way in which Spinoza's thought fits awkwardly with our contemporary paradigm for which discussion of consciousness focuses on discrete entities or "selves." As a result, Spinoza's usage unavoidably raises questions about what it is that "thinks" these "ideas." Spinoza's answer—that these ideas exist in the Mind of God—is not very helpful when we recall that, for him, God is simply the totality of Nature, rather than the anthropomorphized "person" of conventional religion. Nor is the question made any less vexed by the realization that he considers each of these ideas to be "minds" in their own rights: "minds" which have "ideas" of their own.

In the face of such a question, it is helpful to refer back to the interpretations of Bennett,[25] Curley,[26] and Garrett,[27] cited above: first, that Spinoza's conception of

[24] At the level of the totality of nature, of course, the equivalent of *conatus* is not a striving to persist in being, but simply Being (*E*1p20), the idea of which is the attribute of Thought or Mind (*E*2p1).

[25] Bennett 1984.

[26] Curley 1969.

[27] Garrett 2002.

Mind—the Mind of Nature as a whole—is a logical implication of the fact that human beings have minds, and, second, that what Spinoza means by Mind is something like the set of all possible true propositions about extended natural phenomena: the "sense" that nature "makes." Thus, the "minds" of nonhuman "bodies"—be they stones, ecologies, or solar systems—contain "ideas" to the extent that the "sense" that these complex extended phenomena "make" necessarily includes the intelligibility of each of their parts.

2.2.3 Human Psychology: Cognition and Emotion

For our purposes, what is most important about Spinoza's metaphysical conceptions is not really the question of, to paraphrase Thomas Nagle, "what it is like to be a rock,"[28] but rather, how these principles apply to the problems posed by mind/body relations on the human scale; and, in the human context, some of these apparently obscure issues become clearer. The statement that the human mind is the idea of the human body (*E*2p13) still requires explanation, but we are accustomed to the notion that humans have minds and that those minds contain ideas. According to Spinoza, both minds and bodies can be described as active, in the sense of *conatus* or striving to persevere in being. He tells us that the relationship of mind and body, for a given singular thing, is such that the particular body's ability to act (to do things) is matched by the mind's ability to know (*E*3p9&d). Human bodies are extremely complex and capable of doing a great many things in the service of preserving the complex relationships among its parts and between itself and other singular things (including other human beings). And, consciousness arises because the body's ability to do many things in extended space is matched by the mind's ability to know a great many things, including its own ideas, and also itself (*E*2p13s).

For Spinoza, however, true consciousness of self is not, as for Descartes, awareness of self as an independent thinker, but rather, the mind's awareness of itself as a mode of the totality of nature, apprehended under the attribute of thought (*E*5p31s). And his understanding of objects and events other than the self is equally different from that of Descartes or any subsequent positivist epistemology. Seen in its true light, both self and external objects are, for Spinoza, inseparable from Substance and from one another, and any ideas we might have of them, separate from Substance, are inherently only partial and confused, or, in his term, inadequate (*E*2p11c). To him, the experience that Descartes appeals to with the *cogito* is, like all experience and all empirical knowledge, what he calls Knowledge of the First Kind, which, because of its perspectival nature, is inherently flawed. For him, the notion that there is "something it is like to be" anything, a notion that is so central to late twentieth-century debates on consciousness,[29] is of very little interest because

[28] Nagle 1974.

[29] See Chalmers 1996; Dennett 1988; Humphrey 2011; Nagle 1974.

our idea of "what it is like to be" a human being—never mind a bat (or a rock)—is necessarily inadequate.

What is important to Spinoza is not perspectival experience but reason. *Reason* is a term that he uses in two ways. In some cases, reason refers to what I have called "the sense that nature makes," and in other contexts, he uses the term to refer to the step-wise reasoning of deductive processes like mathematics or logic. In the first sense, the one that interests him more, reason is the faculty whereby the human mind, as a Mode of Substance, can grasp top-down causality and can express the eternal order of ideas as they exist in Mind (E2p9c,10c). These ideas, he teaches, can be adequately apprehended either by direct intuition of the essences of singular things or by reasoning (in the second sense) about properties that are common to all singular things. In his cognitive theory, reasoning and intuition are, respectively, the Second and the Third Forms of Knowledge (*E*2p40s2). Though Spinoza values the latter more highly because it is more direct, both provide equally certain answers and result in adequate ideas, which, by definition (*E*2d4), are necessarily true conceptions of a thing as it actually is in Nature. Knowledge of the First Kind can never be adequate because an adequate idea must account for its *ideatum* without reference to anything other than the phenomenon in question, and anything we experience is necessarily a result of networks of bottom-up causality for which we cannot ever fully account. We can know by reason—which is to say by intuition and deduction from either universal laws (*E*2p7) or the eternal essence of Nature (*E*1d3)—that a given simple body's movement or rest is caused by that very body's *conatus*. But, to be adequate, an idea of the proximate causes of the speed and direction of that body's movement would have to include knowledge of an infinite regress of more distal causes, which lie outside our experience. It is only by intuition that we come to an awareness of anything, including ourselves, as a Mode of Substance, the insight with which Spinoza opposes Descartes's experiential *cogito* with its dualist implications.

The fact that most of the ideas we have of ourselves and our environment are based on experience is the key element in Spinoza's theory of affects. He defines affects, or emotions, as "affections of the Body by which the Body's power of acting is increased or diminished, aided or restrained, and at the same time, the ideas of these affections" (*E*3d3). This definition reflects the doctrine that a particular human mind is the idea of a particular body (*E*2p11,13), which has encounters with other bodies in its environment. Awareness of external things is limited to the way they affect our bodies, and, furthermore, he argues that the body itself is known only through the effects of such contacts (*E*2p16). For complex bodies that require external resources, such contacts inevitably have the effect of increasing or decreasing the person's power to continue existing: its *conatus*. His detailed account of the emotions (he lists and defines 47 of them) is built around three primary affects: Desire or Appetite, which is the psychological manifestation of the *conatus*, or striving to exist; Joy, which is the experience of anything that increases the person's power of acting to preserve itself; and Sadness, which is the experience of anything that decreases his or her capacity to act (*E*3p9s, 11s). The other 44 emotions are the combination of one of these primary affects with other ideas—e.g., hope is joy arising from the idea of something in the future (*E*3p18s2).

All emotions entail some element of desire—which as a form of *conatus* is active—and the overwhelming majority are "passions," that is, passive responses to changes in our power to preserve the self resulting from causes external to the person; and therefore, such emotions are understood only inadequately. The only exception is the pleasure we take in reasoning. This pleasure is an exception because reason is the essential activity of the mind; so the pleasure we take in it does not result from any inadequately understood external cause. The person is thus the adequate (complete) cause of this affect (*E*5p3). Most of the time, people act and react on the basis of experience, which is to say on inadequate ideas, and when they do so, they are not really the cause of their own actions, but are passively being acted upon: They are in the grip of a passion. Only when acting from reason can a person be considered a true agent. It is important to note, however, that in Spinoza's theory, because the *conatus* of a human being is ultimately not as powerful as that of the external things that affect it, human reason itself is not sufficient to counteract emotions. It is not reason, but the love of reason or the joy taken in it, that motivates what he calls virtuous behavior (*E*5p7). All human behavior is in some way driven by the desire to preserve the self, but, because the essential activity of the mind is reasoning (*E*3p1,3), the self to be preserved is a reasoning self, which is weakened by the passive affects (passions). The only actions that can successfully contribute to the preservation of the self are those guided by reason in its larger sense, the sense that includes the intuition that it is a mode of Substance (*E*1d5) and the realization that it is subject to the laws of Nature (*E*4apnd32).

2.3 The Reception of Spinoza's Ideas

As noted above, Spinoza's ideas were not well received by his own or the immediately succeeding generations. Because of his rejection of a transcendent creator deity, his contemporaries condemned him as an atheist; and his commitment to the life of reason led him to radically republican political ideas[30] that were considered anathema by nearly every government in Europe.[31] Only in the Dutch Republic under Johan de Witt (1653–1672) were such ideas tolerated; but by the time Spinoza died in 1677, 5 years after de Witt's assassination, conservative Orangist rule in the Netherlands had been restored, and under it, clerical and civil authorities made a concerted effort to prevent publication of his posthumous works. Their attempt failed, but only because of an elaborately planned ruse carried out by Spinoza's literary heirs.[32] Spinoza's works have never gone out of publication since; but official animosity was sufficient to marginalize his philosophy so successfully that for nearly a century after his death it was not safe to discuss his works in print without

[30] de Spinoza 1677/2002.

[31] See Israel 2001; Jacob 2006.

[32] See Gullan-Whur 2000; Nadler 2001.

including an explicit condemnation of them.[33] In the latter part of the eighteenth century, his public reputation would be rehabilitated by the German Romantics, but his new champions were so far removed from Spinoza's scholastic intellectual context that they mistook him for a pantheist,[34] attempting to enlist his work in their reaction against the rationalism of the classical Enlightenment—a rationalism of which he was a principle proponent.

2.3.1 Enlightenment Thought and Research

The result of this effort to marginalize his work was that Spinoza's ideas had very little influence on early research and theorizing about the mind and brain. Early anatomists, in particular, ignored his approach. Willis[35] and (of course) Descartes[36] attempted to explain neurophysiology reductively. They believed the nervous system was composed of hollow tubes through which minute particles (animal spirits) transmitted the impact of external moving particles on sense organs to the brain, which, in turn, directed streams of such particles to activate muscle tissue and understood muscle contraction as resulting from the tissue being inflated by an influx of such particles.

This atomistic, essentially Cartesian, animal spirits model of nerve function persisted well into the eighteenth century, during which it served as the basic model for neuroanatomists like François Pourfour du Petit (1664–1741) and Robert Whytt (1714–1766). In the middle and later parts of the century, the animal spirits accounts would be challenged by anatomists like David Hartley (1705–1757) and Charles Bonnet (1720–1793), who argued that nerves transmit information about the external world by means of vibrations,[37] and by Albrecht von Haller (1708–1777), who opposed the idea that muscles contracted when inflated by animal spirits with the notion that all animal tissue was somehow irritable.[38] But, such alternatives to the animal spirits model, while reflecting more accurate anatomical observation, were no less reductive than the Cartesian model. Bonnet's notion of mind/body relations does seem to have borrowed from Leibniz,[39] who had attempted to reframe Spinoza's holism in a theistic and politically conservative framework, but Spinozan influence seems otherwise missing among these early anatomists.

There are, of course, methodological reasons why natural philosophers engaged in anatomical research would gravitate toward a reductive approach. Their work, unlike Spinoza's, was empirically based, and thus, they were inclined to believe that

[33] Israel 2001.

[34] See DeCuzzani 1991; Goetschel 2004; Jacobi et al. 1916/2010.

[35] Willis 1683/1971.

[36] Descartes 1996b.

[37] See Bonnet 1760/2011; Hartley 1749/1834; Whitaker and Turgeon 2007.

[38] Frixione 2007; von Haller 1735/1936.

[39] Stewart 2006.

the structures they were observing at a micro level were the causes of the more familiar gross behavioral phenomena they sought to explain. Such methodological motives, however, cannot explain the formulations of thinkers like John Locke (1632–1704) and David Hume (1711–1776), whose interest was in the study of human cognition by direct observation alone.

Locke[40] was less thoroughgoing in this endeavor. In spite of expressed intentions to avoid questions about how the mind receives information from the external world (*EHU* Intro.2.), his thinking was clearly influenced by the atomistic animal spirits doctrine that ideas are derived from motion of particles in the external world, producing sensation by causing motion of particles carried to the brain through hollow nerves (*EHU* 2.8.11-13). His theory of ideas—according to which our common understanding of the world is a mental construction created by combining these fundamental and irreducible simple ideas (*EHU* 2.2.1-2)—is, thus, as reductive and atomistic as Descartes's.

Hume[41] was more consistent in his effort to conduct his analysis of the way ideas interact with one another without making any hypotheses about the relation between ideas and their possible causes in the external world (*T.* 1.3.5.2). His reflections led him to an account of mental phenomena in which the association of ideas is grounded only in feeling and habit and cannot be justified by reason (*T.* 1.4.1.12). This contention that thought could not be separated from sentiment or feeling, while not explicitly Spinozan, bears some similarity to Spinoza's affects doctrine (*E*4p14). Also, though he accepted the atomistic "theory of ideas,"[42] his discovery of a logically unbridgeable gap between reason and customary association of ideas showed the limits of the Lockean approach,[43] and confirms Spinoza's designation of empirical knowledge as "inadequate" (*E*2p11c). Furthermore, he is quite Spinozan in his solution to the conflict between reason and habitual thought: Seeing both as products of nature. These Spinozan elements in his thought are probably not accidental; in his youth, Hume was friendly with a group of British Spinozaists, and a number of scholars have argued that he retained more sympathy for Spinoza than he could safely express.[44]

The same is not true of Hume's contemporary, David Hartley, who proposed a somewhat different version of associationist psychology. He, like Bonnet, argued that ideas are formed when vibrations from the external world are transmitted to the brain through the sense organs and along the nerves.[45] This model appealed to Hartley because of its coherence with ideas about physics proposed by Isaac Newton in his *Optiks*,[46] and because by the middle of the eighteenth century, it was known that nerves are not the hollow tubules described in the earlier animal spirits theories.

---

[40] Locke 1690/1959, hereafter "*EHU*".

[41] Hume 1639–1640/2007, hereafter "*T*".

[42] Stroud 1977/2004.

[43] Meehan 2010.

[44] See Giancotti 1999; Klever 1993; Popkin 1979; Russell 1993.

[45] Hartley 1749/1834.

[46] Newton 1730/1952.

The Newtonian element was the idea that the nervous vibrations were supposed to be the same kind of phenomenon as that which allowed gravitational forces to operate at a distance. In this way, Hartley's theory implies a harmonically vibrating universe: An idea which has a certain totalizing thrust. But Hartley, who was a religious man, was unlikely to have been influenced by, or to have even read, Spinoza.

Comparing Hartley's theory of mental associations with that of Hume, it is historically important to note that the former's efforts to tie his account of mental activity to an overly simplified model of physics makes his psychological associationism considerably less flexible than Hume's.[47] This is of interest because it was Hartley's version of associationism that would later be adopted by Jeremy Bentham (1748–1832), whose utilitarian philosophy was a major influence on twentieth-century learning theory.[48] The reception of Hume's model, on the other hand, was limited by the focus of his contemporaries—particularly Thomas Reid (1710–1796) and Immanuel Kant (1724–1804)—on what they erroneously took to be his skepticism.[49]

2.3.2 Nineteenth-Century Research

Given the consummately rationalist nature of Spinoza's work, it is ironic that his public rehabilitation, after a century of censorship by the Moderate Enlightenment,[50] should fall to the antirationalist German Romantics; and this irony is compounded by the fact, noted above, that his new champions, lacking a familiarity with the scholastic roots of Spinoza's *Deus sive Natura* formulation, mistook his holism for pantheism. In spite of such errors, however, the coming of Romanticism marks an important shift in the way European intellectuals approached the study of nature in general and the mind and brain sciences in particular: Before the Romantic reaction, the dominant model had been that of "natural philosophy" with figures from Descartes to Newton and Hartley combining speculative theorizing with experimental research. Afterwards, there was a growing tendency for speculation and experimentation to drift apart so that, by the end of the nineteenth century, science and philosophy would be seen as two separate fields of endeavor.

The career of the Romantic Idealist, Friedrich Schelling (1775–1854), predated this transition. Schelling wanted his *Naturphilosophie* to influence natural science and he devoted considerable effort to experimental work, which he attempted to synthesize into a holistic paradigm, explicitly inspired by Spinoza, in which nature was seen as a kind of organism.[51] Schelling's influence, however, was limited. The

---

[47] Meehan 2011.

[48] Steintrager and Elkins 1977.

[49] Piper 1978–1979; Wood 1990.

[50] Israel 2001.

[51] Richards 2002.

pioneers of what we might call modern neuroscience rejected the vitalist and epigenetic hypotheses to which Schelling's work gave rise. Though certainly aware of Schelling—and often recognizing Spinoza as having foreshadowed some of their findings[52]—they assumed a Cartesian-like conception of bottom-up causality according to which higher-level structures are determined by simpler, more fundamental ones.

This extremely fruitful, early neurophysiological work was, perhaps, best exemplified by the theories and research of Johannes Müller (1808–1858) and his students, who focused detailed investigation of discrete neurological phenomena at the level of cells and reflexes. Their findings include (a) Müller's own formulation of the law of specific energies, (b) the isolation of the cells composing the glia surrounding peripheral nerves by Theodor Schwann (1810–1882), (c) Emil du Bois-Reymond's (1818–1896) discovery of nerve action potential, and (d) Herman von Helmholtz's (1821–1894) measurement of the speed at which electrical signals travel along nerve fibers.[53] These were discoveries that advanced our understanding of nerve function far beyond the animal spirits and vibratory notions of the previous century and provided a basis for much later work, including the psychological research of Wilhelm Wundt (1832–1920). Helmholtz's work in particular was essential for the modernization of the originally Cartesian conception of the discrete reflex,[54] which, in the late nineteenth and early twentieth centuries, would emerge as the primary building block for Ivan Pavlov's (1849–1936) ideas about cortical associationism, B. F. Skinner's (1904–1990) behaviorism, and even the earliest formulations of Sigmund Freud's (1856–1939) *Project for a Scientific Psychology*.[55]

Though these theories were based on the bottom-up causality so central to the Cartesian perspective, some of the important developments of the period did point in the direction of more encompassing formulations, much of which consolidated the earlier micro-level research. This would include: Santiago Ramón y Cajal's (1852–1934) neuron doctrine, also such brain localization research as the cortical mapping work of Korbinian Brodmann (1868–1918) and the discovery of specific language regions by Pierre Broca (1824–1880) and Karl Wernicke (1848–1905). Yet another example is Charles Sherrington's (1857–1952) investigation of the coordination of spinal reflexes into fundamental units of action.[56] The most important integrative work of the period, however, was that of John Hughlings Jackson (1835–1911), who, influenced at least indirectly by Darwin, developed an evolutionary hierarchal model of the brain.[57] Hughlings Jackson's work, in turn, has influenced a variety of later holistically inclined researchers,[58] including Nikolai Bernstein (1896–1966) in Russia and Paul MacLean (1913–2007) in the United States.

[52] Damasio 2003.

[53] Finger 2001.

[54] Platt and Glimcher 1999.

[55] Freud 1950/1966.

[56] Finger 2001.

[57] See Foerster 1936; Smith 1982.

[58] LeDoux 1998.

2.3.3 Spinozan Concepts in Recent Mind/Brain Research

One of the contradictions in twentieth-century neuroscience is that, despite a general acceptance of Cartesian reductive methods, researchers have evidenced a marked distaste for Descartes's mind/body dualism largely because it implies that human thought and experience lie outside the natural world and beyond the reaches of scientific methods.[59] Some, like David Chalmers[60] and Thomas Nagle,[61] have argued that, because it is perspectival, human experience is not analyzable in objectively scientific terms: hence, what is referred to in the literature as "the hard problem." However, the argument of Patricia Churchland's *Neurophilosophy*,[62] that human thought could ultimately be explained in neurological terms, was more widely accepted.[63] Yet, as our understanding of neuroscience has improved, this Cartesian method seems to be reaching the limits of its usefulness, and some researchers, particularly those studying movement and the emotions, have begun to consider Spinoza-like alternatives to a forced choice between dualist and reductive-materialist solutions to "the hard problem."

By and large, contemporary researchers who pursue these Spinoza-like conceptions have arrived at them independently and make little or no direct reference to Spinoza. One exception to this generalization is Antonio Damasio, whose *Looking for Spinoza*[64] is an explicit effort to relate Spinoza's philosophy to contemporary neuroscience. At the heart of Damasio's argument in this book are two Spinozan ideas: the notion that the mind is the idea of the body (*E*2p13), and the assertion, which pervades the Third Part of the *Ethics*, that this mind/body unity implies the centrality of emotion in cognitive processes. For Damasio, the mind is a collection or stream of the mental images of bodily events. These events are responses to what he calls "emotionally competent stimuli,"[65] and in an earlier work,[66] he presented research evidence that damage to emotion processing structures in the brain impairs decision-making processes in ways that compromise an individual's ability to survive. Absent the neurological images of bodily responses to environmental stimuli, an organism has an impaired sense of self and cannot, in Spinoza's terms, "endeavor to persist in being" (*E*3p6).

Damasio's model also parallels Spinoza's idea that the *conatus* of singular things is manifest, not only in the momentum exhibited by simple objects, but in the dynamic internal relations and homeostasis of complex ones (*E*2a2'd). For Damasio, the emotions and feelings of complex organisms are the product of a nested

[59] See Damasio 1994; Humphrey 2011.

[60] Chalmers 1996.

[61] Nagle 1974.

[62] Churchland 1982.

[63] E.g., Dennett 1991; Freeman 2000.

[64] Damasio 2003.

[65] Ibid., p. 53.

[66] Damasio 1994.

hierarchy of reactions in which, for example, metabolic processes are elements of immune reflexes, and both are mobilized in pain and pleasure behaviors, drives, and complex emotions like love and sympathy.

Spinozan as Damasio's work is, he still preserves a somewhat bottom-up approach that results in a more passive conception of the mind than Spinoza proposed. Damasio's main focus is on presenting a theory of emotions, a modernized version of the James-Lange theory, and while he does an excellent job of showing how his findings were foreshadowed by many of Spinoza's ideas about mind/body relations, he does not seem to capture the full antireductionist import of Spinozan metaphysics. There is a way in which Damasio's nested hierarchy is constructed along the lines of bottom-up causality with higher functions seen as complex reflexes. Furthermore, while he conveys Spinoza's understanding of the ways our emotions are driven by environmental events and organized around the impulse to survive, he misses the insights about the *human* mind as rational and motivated to preserve itself as a rational entity. In Damasio's model, the mind receives emotionally significant information and reacts to it, but, while Damasio is certainly aware of the active role of attention in emotional/cognitive processes, he does not convey Spinoza's vision of the mind as striving to preserve and foster its own rational capacities motivated by an intellectual love of God—which is to say, Nature (*E*5p32c).

More closely aligned with Spinoza's notion of an intellectual emotion is research reported by Jaak Panksepp in his *Affective Neuroscience: The Foundations of Human and Animal Emotions*.[67] Panksepp, unlike Damasio, makes no explicit reference to Spinoza but his work—building on Olds and Milner's[68] neuroanatomical research into self-stimulation and on Edward Chase Tolman's[69] studies of animal behavior and motivation—suggests an understanding of human and animal minds as less dependent on external stimuli and more active than that depicted by Damasio. Specifically, Panksepp focuses on a neural system, not described by Damasio, in the lateral hypothalamic corridor, which supports an emotional urge to explore the environment for its own sake rather than as a means of obtaining satisfaction for drives related to immediate survival or reproduction. The behaviors, as well as the hormonal and sympathetic nervous activity evidenced when this system is active, are quite dissimilar to those that accompany the animal's consumption of a conventional "reward." In Panksepp's model, the self-activating "seeking system" provides a neurological and affective basis for the animal's sense of self, and in humans, a sense of self-conscious ego. This, of course, is precisely the kind of neuropsychological structure we would expect to find in a being that Spinoza describes as needing to understand the world of which it is a part (*E*5p7).

In addition to Damasio's work on the interaction of emotion, mind, and body—and Panksepp's understanding of a primary-seeking drive: curiosity—neuroscientists are also discovering phenomena and processes for which they can account only

[67] Panksepp 1998.

[68] Olds and Milner (1954).

[69] Tolman 1922/1958.

in terms of a Spinoza-like, if not explicitly Spinozaist, holism with its recognition of top-down causality. The most profound impulse towards biological holism was, of course, the development of evolutionary theory, completed by Charles Darwin (1809–1882) in his *Origin of Species*.[70] The evolutionary approach was applied specifically to neuroscience by John Hughlings Jackson, following the alternate model of evolution proposed by Herbert Spencer (1820–1903).[71] The key premise of any post-Darwinian evolutionary theory is the interplay of variation and selection: with specific, isolated (and thus atomistic) mutations exercising a bottom-up causality while nature as a whole has a top-down causal effect through the process of selection.[72]

Hughlings Jackson's importance to the current narrative, however, is not limited to his evolutionary theorizing. His interest in epilepsy led him to the study of motor activity, and through this aspect of his work, as noted above, he influenced Nikolai Bernstein's studies of the motor system. These studies moved beyond the Cartesian assumptions of Sherrington's analysis of the way reflexes are integrated into fundamental units of action in the spinal cord,[73] to the study of how the whole organism exercised top-down control, limiting and directing the vast range of possible movement to those specific actions needed for its immediate purpose (Bernstein 1998). Bernstein also elaborates a very Spinozan notion of the interrelationships of mind, body, and environment in his observation that the body, in particular the hand, represents objects in the environment by virtue of the ways in which it adjusts itself to conform to those objects. In his account, these bodily "representations" are the physical correlates of a mental activity, such that dexterity is a kind of intelligence.

This notion, of course, is a direct contradiction of the Cartesian mind/body dualism that gives rise to "the hard problem." It also suggests that what we think of as the human mind might not be wholly located in the person: It might, to some degree, be distributed or situated in the environment. This distributed intelligence model has been elaborated by Andy Clark[74] and others.[75] In such a model, the mind is situated in both the individual and the environment, which, though stated without the language of seventeenth-century scholastic rationalism, is suggestive of Spinoza's teaching that the human mind, far from being a discrete entity, is a mode of the totality of which it is a part (*E*2p5&11c).

This holistic notion of top-down causality, combined with a Spinoza-like emphasis on the mind/body (organism) as actively striving to persevere in being, is also implicit in the late twentieth-century focus on the relation between motivation and action. One researcher, Charles Gallistel,[76] defines motivation in terms of the potentiation/depotentiation of motor subroutines and traces connections between

[70] Darwin 1859/2003.

[71] Smith 1982.

[72] See Monod 1971; Polanyi 1967.

[73] Sherrington 1961.

[74] Clark 1997.

[75] For a comprehensive review, see Chandrasekharan and Osbeck 2010.

[76] Gallistel 1980, 1993.

high-level control systems and lower-level neurobehavior: work that can be seen as an extension and verification of Bernstein's. Another contemporary neuroscientist, Marc Jeannerod,[77] presents evidence supporting a simulation theory according to which the mind creates images of external objects in terms of the neural patterns necessary to interact with those objects: what he calls action images. Ralph Ellis and Naktia Newton[78] have further argued that the mind's repertoire of action images includes not only the neural traces of actualized interactions with the external world, but also the traces of motor commands that were inhibited before they could be activated. In any coordinated action images, the order of the neural traces necessarily matches the order of the actualized or imagined physical motions; this notion comes very close to Spinoza's doctrine that the order of ideas in the mind matches the order of physical things (*E*2p7). Such action images are, in Cartesian fashion, grounded in the cellular activity of the neurons,[79] but are also shaped in top-down fashion by the organism's larger purposes and the structure of the environment in which that organism can imagine acting out its purposes.

2.4 Conclusion

As my title suggests, the narrative of Spinozan concepts in the mind and brain sciences is, in many ways, a story of the return of the repressed: ideas censored and ignored in most of early-modern natural philosophy and in the early stages of modern neuroscience being rediscovered and reintroduced, either explicitly or implicitly, after the political and religious conditions that gave rise to the censorship had passed and when the problems inherent in the Cartesian paradigm began to outweigh its usefulness. These ideas include the notion of top-down causality: a conception of mind as the representation of the body (c.f., *E*2p13) based on its interaction with the environment (c.f., *E*2p16) and an understanding of human nature, in both its physical and mental abilities, as an integral part and product of nature as a whole (c.f., *E*2p9c,10c).

The suitability of Spinoza-like ideas for overcoming the limitations of Cartesian reductionism and mind/body dualism is, of course, to be expected, given that both Spinoza's metaphysics and his biology were developed as explicit refutations of Descartes. But the reemergence of these approaches is not exactly a "rediscovery": Intimations of holism can be found in Hume, and in a somewhat distorted form, Spinoza's ideas were asserted by the German Romantics and thus kept alive as at least a possible approach for nineteenth-century biology and neurophysiology. The key to the resurgence of Spinozan ideas in modern mind and brain sciences, however, was Darwin's theory of evolution, which, as a secular explanation of nature as a whole that treats organisms as the products of a self-caused—or at least

[77] Jeannerod 1997.

[78] Ellis and Newton (2010).

[79] Lethin 2002.

self-organizing—nature (cf., *E*1d1), refutes the Cartesian *cogito*. When nature is understood as evolutionary process, Descartes's conception of the self as a discrete entity becomes incoherent, as does the notion that the mind and body are anything other than products of the same process that has shaped the one along with the other (cf., *E*2p10). And, when the behavior of organisms is seen to be motivated by survival needs (cf., *E*3p6), sharp distinctions between thought and emotion become equally meaningless (cf., *E*3Preface).

The totalizing effect of evolutionary theory is, thus, comparable to Spinoza's holism, allowing a change whereby the cell and reflex-level neuroscience could be augmented by the whole brain approaches of Hughlings Jackson and Bernstein, among others. So, too, Darwin's emphasis on survival and adaptation, which parallels Spinoza's *conatus*, are the key elements in both Damasio's and Panksepp's understanding of motivation and emotion. These elements can also be found in the work of Gallistel, Jeannerod, and others in which mental imagery and cognition are understood in terms of intentional interaction with the environment.

References

Allison HE (1975) Benedict de Spinoza. Twayne Publishers, Boston

Aquinas T (1270) Summa theologicae. Retrieved from http://www.fordham.edu/halsall/source/aquinas3.html

Bennett JF (1984) A study of Spinoza's ethics. Hackett, Indianapolis

Bernstein N (1998) Dexterity and its development (resources for ecological psychology). Psychology Press, Florence

Bonnet C (2011) Essai analytique sur les facultés de l'âme [Analytic essay on the powers of the soul]. N.P.: Books on Demand (Original work published 1760)

Chalmers D (1996) Facing up to the problem of consciousness. In: Hameroff SR, Kaszniak AW (eds) Toward a science of consciousness: the first Tucson discussions and debates. Complex adaptive systems. MIT Press, Cambridge, MA, pp 5–28

Chandrasekharan S, Osbeck L (2010) Rethinking situatedness: environment structure in the time of the common code. Theory Psychol 20:171, http://tap.sagepub.com/cgi/content/abstract/20/2/171

Churchland PS (1982) Neurophilosophy: towards a unified science of the mind/brain. MIT Press, Cambridge, MA

Clark A (1997) Being there: putting brain, body, and world together again. Bradford Books, Cambridge, MA

Curley EM (1969) Spinoza's metaphysics: an essay in interpretation. Harvard University Press, Cambridge, MA

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

Damasio A (2003) Looking for Spinoza: joy, sorrow, and the feeling brain. Harcourt, Orlando

Darwin C (2003) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. Signet Classics, New York (Original worked published 1859)

de Spinoza B (1985) Ethics. In: Curley E (ed & trans) The collected works of Spinoza, vol I. Princeton University Press, Princeton, pp 408–617 (Original work published 1677)

de Spinoza B (2002) Political treatise. In: Shirley S (trans), Morgan ML (ed) Complete works. Hackett Publications, Indianapolis, pp 680–754 (Original work published 1677)

DeCuzzani P (1991) Spinoza et les Spinozismes: de Oldenburg à Hegel, L'histoire d'une repudiation [Spinoza and the Spinozisms: from Oldenburg to Hegel, the history of a repudiation]. Revue d'Histoire et de Philosophies Religieuses 71:349–364
Dennett D (1988) Quining qualia. In: Marcel AJ, Bisiach E (eds) Consciousness in contemporary science. Oxford University Press, New York, pp 42–77
Dennett D (1991) Consciousness explained. Little Brown, Boston
Descartes R (1996a) Discours de la method. In: Adam C, Tannery P (eds) Oeuvres de Descartes, vol 6. Librarie Philosophique J. Vrin, Paris, pp 1–78 (Original work published 1637)
Descartes R (1996b) Les passions de l'âme. In: Adam C, Tannery P (eds) Oeuvres de Descartes, vol 10. Librarie Philosophique J. Vrin, Paris, pp 327–497 (Original work published 1649)
Descartes R (1996c) Le monde, où traite de la lumière. In: Adam C, Tannery P (eds) Oeuvres de Descartes, vol 11. Librarie Philosophique J. Vrin, Paris, pp 3–202 (Original work published in 1662)
Ellis RD, Newton N (2010) How the mind uses the brain to move the body and image the universe. Open Court, Chicago
Finger S (2001) Origins of neuroscience: explorations into brain function. Oxford University Press, New York
Foerster O (1936) The motor cortex in man in the light of Hughlings Jackson's doctrine. Brain 59(2):135–159
Freeman WJ (2000) How brains make up their minds. Columbia University Press, New York
Freud S (1966) Project for a scientific psychology (trans: Stachey J). In Strachey J (ed) The standard edition of the complete psychological works of Sigmund Freud, vol 1. Hogarth Press, London, pp 283–346 (Original work published 1950)
Frixione E (2007) Irritable glue: the Haller-Whytt controversy on the mechanism of muscle contraction. In: Whitaker H, Smith CUM, Finger S (eds) Brain, mind, and medicine: essays in eighteenth century neuroscience. Springer, New York, pp 115–124
Gallistel CR (1980) From muscles to motivation. Am Sci 68(4):398–409
Gallistel CR (1993) The organization of learning (learning, development and conceptual change). Bradford Books, Cambridge, MA
Garrett D (2002) Spinoza's *conatus* argument. In: Koistinen O, Biro J (eds) Spinoza: metaphysical themes. Oxford University, Oxford, pp 127–158
Gassendi P (1972) The selected works of Pierre Gassendi. Brush CB (ed and trans). Johnson Reprint Corp., New York
Gaukroger S (1995) Descartes: an intellectual biography. Oxford University Press, New York
Giancotti E (1999) The theory of the affects in the strategy of Spinoza's ethics. In: Yirmiyahu Y (ed) Desire and affect: Spinoza as psychologist [Spinoza by 2000, the Jerusalem conferences: Ethica III]. Little Room Press, New York, pp 129–138
Goetschel W (2004) Spinoza's modernity: Mendelssohn, Lessing, and Heine. University of Wisconsin Press, Madison
Gullan-Whur M (2000) Within reason: a life of Spinoza. St. Martin's Press, New York
Hartley D (1834) Observations on man, vols 1 and 2. Thomas Tegg and Son, London. http://www.archive.org/stream/observationsonma01hartuoft (Original work published 1749)
Hume D (2007) A treatise of human nature: vol. 1. In: Norton DF, Norton MJ (eds) Texts. Oxford University Press, Oxford (Original work published 1739–40)
Humphrey N (2011) Soul dust the magic of consciousness. Princeton University Press, Princeton
Israel JI (2001) Radical enlightenment: philosophy and the making of modernity 1650–1750. Oxford University Press, Oxford
Jacob MC (2006) The radical enlightenment: pantheists, freemasons, and republicans, 2nd edn. Cornerstone Book Publishers, Lafayette
Jacobi FH, Mendelssohn M, Scholz H (2010) Die Hauptschriften Zum Pantheismusstreit Zwischen Jacobi Und Mendelssohn. Nabu Press, Charleston (Original work published 1916)
Jeannerod M (1997) The cognitive neuroscience of action. Blackwell, Oxford
Klever W (1993) More about Hume's debt to Spinoza. Hume Stud XIX(1):55–74

LeDoux J (1998) The emotional brain: the mysterious underpinnings of emotional life. Simon & Schuster, New York
Lethin A (2002) How do we embody intentionality? J Conscious Stud 9(8):36–44
Locke J (1959) An essay concerning human understanding (complete and unabridged; collated and annotated by A. C. Fraser; vols. 1–2). Dover Publications, New York (Original work published 1690)
Meehan W (2009) *Partem Totius Naturae Esse*: Spinoza's alternative to the mutual incomprehension of physicalism and mentalism in psychology. J Theor Philos Psychol 29(1):47–59
Meehan W (2010) Foreshadowing natural selection: Hume and the limit of Locke's Atomism [Poster]. Paper presented at the American Psychological Association, San Diego, CA. Available at http://wmeehan.net/downloads/foreshadowing_natural_selection.pdf
Meehan W (2011) Psychology in the age of Newton: Hume and Hartley on the association of ideas [Poster]. Paper presented at the American Psychological Association, Washington, DC. Available at http://wmeehan.net/documents/PsychologyintheAgeofNewton.pdf
Monod J (1971) Chance and necessity: an essay on the natural philosophy of modern biology. Alfred Knopf, New York
Nadler S (2001) Spinoza: a life. Cambridge University Press, Cambridge
Nadler S (2011) Spinoza the atheist. New Hum. Retrieved from http://newhumanist.org.uk/964/spinoza-the-atheist
Nagle T (1974) What is it like to be a bat? Philos Rev 83:435–450
Newton I (1952) Optiks, or a treatise of the reflections, etc. [Foreword by Einstein A; Introduction by Whittaker E; Preface by Cohen IB, and analytical table of contents by Roller D]. Dover Publications, Mineola (Original work published 1730)
Olds J, Milner P (1954) Positive reinforcement produced by electrical stimulation of the septal area and other regions of rat brain. J Comp Physiol Psychol 47:419–427
Panksepp J (1998) Affective neuroscience: the foundations of human and animal emotions. Oxford University Press, Oxford
Piper WB (1978–1979) Kant's contact with British Empiricism. Eighteenth-Century Stud 12(2):174–189
Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. Nature 400(6741):233–238
Polanyi M (1967) The tacit dimension. Doubleday Books, New York
Popkin RH (1979) Hume and Spinoza. Hume Stud 5(2):65–93
Ravven HM (2003) Spinoza's anticipation of contemporary neuroscience. Conscious Emotion 4:257–290
Richards RJ (2002) The romantic conception of life: science and philosophy in the age of Goethe. University of Chicago Press, Chicago
Russell P (1993) Epigram, pantheists, and free-thought in Hume's treatise: a study in esoteric communication. J Hist Ideas 54:659–673
Sherrington CS (1961) The integrative action of the nervous system. Yale University Press, New Haven
Smith CUM (1982) Evolution and the problem of mind: Part II. John Hughlings Jackson. J Hist Biol 15:241–262
Steintrager J, Elkins J (1977) Bentham. Cornell University Press, Ithaca
Stewart M (2006) The courtier and the heretic: Leibniz, Spinoza, and the fate of God in the modern world. W. W. Norton, New York
Stroud B (2004) Hume. Routledge, London (Original work published 1977)
Tolman EC (1958) A new formula for behaviorism. In: Behavior and psychological man: essays in motivation and learning. University of California Press, Berkeley, pp 1–8 (Original work published 1922)
von Haller A (1936) A dissertation on the sensible and irritable parts of animals (translated from the Latin). Johns Hopkins Press, Baltimore (Original work published 1735)

Whitaker HA, Turgeon Y (2007) Charles Bonnet's neurophilosophy. In: Whitaker HA, Smith CUM, Finger S (eds) Brain, mind and medicine: essays in 18th-century neuroscience. Springer, Norwalk, pp 191–200

Willis T (1971) Two discourses concerning the soul of brutes, which is that of the vital and sensitive of man [Facsimile reproduction of the translation by S. Pordage]. Scholars Facsimiles and Reprints, Gainesville (Original work published 1683)

Wood PB (1990) The natural history of man in the Scottish enlightenment. Hist Sci 28(1):89–123

Chapter 3
'Struck, As It Were, with Madness': Phenomenology and Animal Spirits in the Neuropathology of Thomas Willis

Kathryn Tabb

> *Wherefore to explicate the uses of the Brain, seems as difficult a task as to paint the Soul, of which it is commonly said, That it understands all things but itself (Preface to "The Anatomy of the Brain"* (Willis 1681a)*).*

3.1 Willis's Felt Neuropathology

The embrace of neuroscience by psychiatric medicine over the past decades has been praised and damned – heralded as evidence that psychiatry has taken its rightful place among the sciences, and reviled as an abandonment of the psychological, experiential, and spiritual aspects of mental illness. The biomedical paradigm in psychiatry has reconceived the patient as a dysfunctional organism, and aims to identify and intervene upon pathological mechanisms within the brain. As a result, psychoanalysis, along with derivative flavors of psychotherapy based around personal narrative and experience, has been marginalized, despite its dominance for much of the twentieth century. Biomedical psychiatry seems to annex organic neurological dysfunction for itself while denying the hard problems of mental illness, or assigning them to other, softer sciences and disciplines: psychology, social work, philosophy, and religious counseling. The assumption of this reductive approach is that the problem of pathological consciousness will slowly dissolve in a properly neuroscientific solvent.

Many trace the origins of this reductive turn to Thomas Willis,[1] the seventeenth-century doctor and founding member of the Royal Society whose innovations

[1] Especially neuroscientists – for examples see Ochs (2004), Williams (2002), or Wallace (2003) – as well as historians of medicine (for an index of these see Rousseau [2004, p. 181, n. 19]).

K. Tabb (✉)
Department of History and Philosophy of Science, University of Pittsburgh,
1017 Cathedral of Learning, 4200 Fifth Avenue, PA 15260, Pittsburgh
e-mail: Kct5@pitt.edu

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_3,

included the coinage of the term "neurology," the localization of individual mental functions within the brain, and etiological accounts of psychopathology that seem to many to be prescient. As a figurehead of the seventeenth-century sciences of mind, Willis elicits polarized responses from modern-day commentators. For some, he is the arch-reductionist who regrettably laid down "the foundations of a psychiatry without psychology" (Zilboorg and Henry 1941, pp. 265–266); for others the humanist-physician who "inspired a revolution in intellectual thought concerning the nature of man" (Rousseau 2004, p. 166). Through an examination of a key component of Willis's neuropathology – his evasive and pervasive *animal spirits* – I argue that neither of these readings is, on its own, satisfactory.

Even as Willis presented highly original explanations for mental illness based on his iatrochemical corpuscularism that seem compatible with the larger mechanistic projects advanced in his day, his psychopathology maintained a phenomenological core. I will suggest that this is due to the vitalist strain in his iatrochemistry which, rooted in the alchemical tradition, gave to matter a variety of powers out of which life, and mind, were generated. In cases of gross as well as more subtle brain injury, Willis's descriptions of mental disorder often do not reduce the telltale symptoms – sadness, fury, stupidity, hysteria – to matter in motion. Rather they are relocated, attributed not only to the patient but also to the myriad of minute agents that course through his nerves, either with "regular motion" or "inordinate" leaps, some "tender," some "strong," some "unquiet and furious," some "evilly disposed": the animal spirits (Willis 1681b).[2]

Despite the temptation to ignore what has deservedly been called Willis's "puzzling physiology"[3] (Sutton 1998, p. 131) a genealogy of the problem of consciousness in philosophy of mind should take into account how deeply Willis embedded the subjective component of nervous disorders into his theories. Willis's neurological accounts are multi-leveled and pluralistic rather than reductive insofar as both the composition of matter and its vital and psychic powers play their parts. As Lester King has noted, careful attention to Willis rewards with a reminder of "the futility of pursuing any schema of paradigms" (King 1978, p. 143) in the history of early modern natural philosophy. While Willis certainly made great and original strides by theorizing about the organic bases for mental disorder, he cannot, I will argue, be justly accused of (or lauded for) urging the reductive

[2] My project of untangling vitalization from anthropomorphization in Willis's account of the animal spirits is made more delicate by the shifting significance of the components of the term: today our contrast class for "animal" is "human" or "inorganic," rather than "vital" or "natural." Recalling that the root of the word is from "soul" (*anima*) is helpful here. Similarly, Rousseau (2004, p. 20) claims of the medieval concepts of animal and Holy spirit that "each was embodied in the other, if not an extension of itself then a mirror reflection." While this description may also aptly apply to the contemporary reader's associations with the term "spirit," it should be recalled that Willis was deeply saturated in the iatrochemical tradition in which fermentation and distillation were fundamental forces. "Spirit" was a technical term of art.

[3] Another temptation is to laud Willis for what he got "right" and dismiss the rest – see Rather (1974) and Eadie (2003a, b). I side with Frank (1990) that "[Willis's] ideas are 'wrong' in ways which, to me as a historian, are highly appealing."

turn that has found its acme in the biomedical paradigm. His ontology is more complicated than the insistently mechanistic materialism of the contemporary biomedical psychiatrist.

3.2 Willis's Two Souls

In *Two Discourses Concerning the Soul of Brutes* Willis locates himself in a tradition rooted in the Epicurean notion of the "corporeal soul," which recognizes a sensitive and motive force within living things. However, like many of his contemporaries, he also recognizes an additional and very un-Lucretian rational soul alongside the corporeal soul we share with brutes (Wright 1991).[4] Again like many of his contemporaries, Willis observes that the rational soul allows humans to reason and judge; to think in universals where animals cannot, and to abstract from experience; and, perhaps most importantly, to receive divine judgment and receive their eternal reward or punishment.[5] In his discussion of psychopathology, however, Willis does not spare much ink for the rational soul, which he locates in the imagination in the "Middle or Marrowie part of the Brain"[6] (1683, p. 41), rejecting, on anatomical grounds, Descartes's location of it in the pineal gland (Ochs 2004, p. 75).

Rather than figuring in the etiology of mental illness, the rational soul's role is passive – in madness its ability to impose reason on the imagination becomes compromised through bodily changes: "the Rational Soul whilst in the Body, hangs or depends as to its acts and habits, because the Organs being hurt, or hindred, a privation or an Eclipse of these succeeds" (Willis 1683, p. 32). Because all aspects of personality and interpersonal difference are due to the particulars of the sensitive soul[7] rather than the rational one, disorders of the nervous system are all seen as of a

---

[4] In his discussion of the Epicurean inheritance of Locke and Willis, Wright (1991) notes how Willis's location of Descartes in this tradition alongside Gassendi and Digby, while seeming bizarre at first to the contemporary reader, demonstrates that in Willis's milieu "soul" signified precisely those life functions that Descartes banished to the body. Wright emphasizes, however, that Willis's corporeal soul was ultimately far more expansive, bestowing on animals a "'sensitive' use of reason" – a sort of ratiocination (1991, p. 249). Indeed, he goes so far as to suggest that, "For Willis, the higher soul perceives the images of the lower soul and so operates on and reacts to an entity which is already thinking (in Descartes's sense)" (Wright 1991, p. 253).

[5] See Thomson (2008, pp. 81–82) for a discussion of how Willis's Anglicanism may have motivated his insistence on the importance of the rational soul in an (not entirely successful) attempt to preempt accusations of unorthodoxy. See also Kassler (1998).

[6] Otherwise known as the corpus callosum.

[7] Willis distinguishes between the sensitive soul, constituted by the spirits in the nerves and brain, and the vital soul, which plays little role in sensation or perception but is responsible for activities of bodily maintenance like breathing, digestion, and the flow of blood. Nonetheless he refers to man as a two-souled creature, conflating the vital and sensitive functions into one corporeal soul.

kind – Willis considers conditions such as gout and convulsions alongside those seen today as "mental," like depression, mania, and hysteria.

Thus the focus of Willis's psychopathology is not the rational soul but instead the second, corporeal soul made up of the animal spirits and their complex milieu of passions, memories, and imaginings. As Yvette Conry has emphasized, Cartesian metaphysics was not a hospitable framework for psychopathology, as the rational soul was immaterial and could not be injured and the body was, by definition, not mental. With his corporeal soul, capable of sensitive and vital functions and, as Wright (1991) emphasizes, operating as an active agent sometimes in tandem with and sometimes in rebellion against the rational soul, Willis makes mental illness an object that can be studied within the naturalistic framework of the new sciences, especially iatrochemistry (Conry 1978, p. 209).

That said, Willis's publications on the anatomy and pathology of the brain did little to change the clinical treatment of psychopathology in the early modern period (Hunter and Macalpine 1963, p. 189). Indeed, alongside iatrochemical interventions he often explicitly promoted Galenic nostrums that had been in widespread use for centuries and would continue to be the basis for psychopharmacology well past his time. However, he did undergird traditional medicine with a radical new foundation, incorporating not only his own anatomical findings but also the advances of William Harvey and other contemporaries (Rather 1974, pp. 71–112). At the heart of his new science of the brain and nerves was a chemical analysis of the animal spirits.

3.3 Sensions and Spirits

Scholars trace the lineage of the early modern concept of animal spirits back to Aristotle, who theorized spirit (*pneuma*) to be the source of bodily motion. According to Aristotelian theory the heart, which houses the soul, generates *pneuma psychikon* that "appears to stand in soul-origin in a relation analogous to that between the point in a joint which moves being moved and the unmoved" (Aristotle 1984, pp. 703^{a}12–13).[8] In other words, the spirits' *sui generis* power of movement transforms the intentions of the soul into actions by animating the passive matter of the body. Aristotle's concept was expanded on by Erasistratus and, most notably, by Galen: while he relocated the transformation of vital spirit into animal spirit from the heart to the brain, Galen followed Aristotle in considering animal spirits "the first instrument of the soul."

The classical concept, or at least the term *spiritus animalis*, resurfaced in England as early as 1250. The first known English usage – "animal spirits" – was in 1425. A popular medical dictionary, first published in 1688, shows the dominance of the Aristotelian influence but also the modifications that the concept had undergone

[8] See also Chap. 1 of this volume.

by the close of the seventeenth century: "Animal spirits," Blankaart reports, "are very thin fluids which trickle in the cortical substance of the brain from blood and are then raised up into spirits and flow […] into the nerves and the spinal medulla (spinal cord) and illuminate them by accomplishing sensations and actions" (quoted in Jarcho 1982, p. 576). The vagueness of this definition reflects seventeenth-century disputes over the ontological status of the spirits,[9] as well as over the motive and sensitive power that was traditionally attributed to them.[10]

Descartes conceives animal spirits as a subtype of the blood that is uniquely light and quick, able to pass through the less "subtle" matter of the body in order to animate it. On this view bursts of animal spirits released from the brain "get out through the pores in the substance of it; which pores convey them into the nerves, and from thence into the muscles, by means whereof they mould the body into all the several postures it can move" (Descartes 1650, p. 9). By agitating the animal spirits already present in one muscle, animal spirits sent from the brain are able to cause the paired muscle to contract. By controlling the density of animal spirits in each muscle of the pair, the original impulse can generate movement. The number of animal spirits and the violence or sluggishness of their activity are, for Descartes, also key to the etiology of emotion, memory, and insanity.

Willis's account of motion is similar to Descartes's insofar as he believes movement to be generated by the displacement of animal spirits as a result of nervous activity by the brain. However, following Gassendi, Willis argues that animal spirits can catalyze copulas, or sudden outward thrusts of motion, within muscles in order to cause contractions.[11] He also attributes an additional kind of action to animal spirits, in which they relay information in a chain formation through the nerves rather than by individually traveling the length from "organs of sension" to the brain. One of Willis's main preoccupations is the question of how the "many Nerves which serve for the Sense of Feeling, do in a like manner serve for performing the Motions of those Parts to which they belong" (1683, p. 63). He concurs with Descartes that animal spirits flow to the muscles to induce movement, and agrees that the spirits carve tracks in the brain by their passage in order to preserve memory and generate thought. For sensation, however, he suggests that a myriad of animal

[9] Nonetheless I would tend to agree with Sutton (1998, p. 45) that their ontology was of less general interest than their ability to be controlled, both by the rational soul and by the physician.

[10] The broader role of animal spirits in neurology was also challenged as well as championed during the seventeenth century. Harvey was explicitly contemptuous of spirit talk, and Descartes's theory of muscular contraction was subject to particularly incisive attacks by Jan Swammerdam, who demonstrated that a muscle could shorten without the volume of its paired muscle increasing (as it would following an influx of animal spirits). In the following century Robert Whytt argued strenuously against the anthropomorphic nature of animal spirits, arguing that the movement of nervous fluid should be explained by the natural and law-like *telos* of the soul rather than by discrete quasi-mechanistic atoms. See Rocca (2007).

[11] For a synopsis of Gassendi and Willis's joint assault on the Cartesian treatment of the nervous system, see Wallace (2003). For Gassendi's broader influence on Willis see Meyer and Hierons (1968) and Wright (1991).

spirits are implanted in the thick but porous walls of the nerves whose job is to keep their stations – "whether they be set in Battel Array, or on Watch" (Willis 1683, p. 56) – but, through their jiggling movement, to pass on information gathered at the periphery of the body.

Perhaps inspired by his colleague Hooke's work on the nature of light,[12] Willis repeatedly draws an analogy between the excitation of the pores of the air by a ray of light and the excitement of animal spirits by "sensions": "by a continued Series of the Animal Spirits, as it were an Irradiation, the Type of its Impression doth pass from the Sensory to the Head" (Willis 1683, p. 58).[13] By positing a wave-like undulation of information – another favored analogy is the string of a lute – Willis solves the problem of how sensions could move towards the brain without conflicting with outgoing messages from the cerebrum or the cerebellum, the locus for instinctual actions and reflexes.[14] Like multiple waves in a pool, they can, in certain cases, pass through each other intact.

Willis paints a very different picture of animal spirits than the Cartesian: instead of dumb bits of matter generating a hydraulic force to move muscle tissue, they are rather the messengers of the sensitive soul. And in terms of their generation and constitution, Willis's spirits are very different beasts from Descartes's: they are generated through chemical processes rather than by being filtered from the grosser matter of the blood. Willisian spirits, unlike their Cartesian counterparts, are characterized in chemical terms and can undergo chemical transformations. Heavily influenced by the iatrochemical tradition of Paracelsus and van Helmont, Willis believes matter to be made up of seminal principles that have diverse active powers. Following Glisson, he attributes change in organic forms to the fermentation of spirits from states of fixation to fusion and volatility (Clericuzio 1994, pp. 60–61). He explains his preference for "chymistry" over the four principles of Aristotle or various forms of atomism in the following manner: "Because this Hypothesis determinates Bodies into sensible parts, and cuts things open as it were to the life, it pleases us before the rest" (Willis 1684, p. 2). While Descartes says of animal spirits that "no other power propels them than that inclination which they possess to continue their movement according to the laws of nature" (2003, p. 28), Willis criticizes Descartes's "Reasons for the mechanical provision of living Creatures" as unable "to satisfie a Mind desirous of Truth": he argues emphatically that, "on the contrary, Atoms, which are the matter of sublunary things, are so very active and self-moving" (Willis 1683, p. 3). Spirit is an active principle, and the animal spirits represent spirit in its lightest and most perfect form, the result of complex processes of fermentation

[12] See Conry (1978) for a discussion of Hooke's influence on Willis's metaphorical armament.

[13] On "sension:" this favorite word of Willis's refers to the "Symbol," impression or effect caused by a sensed external object, which, passed along by animal spirits through the nerves, is finally projected onto the inner recesses of the brain as if from a camera obscura. The Century Dictionary (1889) defines "sension" as "The becoming aware of being affected from without within sensation."

[14] For more on the role of animal spirits in the relationship between sensation and motion see Canguilhem (1977), especially the third chapter, which treats Willis.

within the blood and the matter of the brain: "Spirits are Substances highly subtil, and Aetherial Particles of a more Divine Breathing, which our Parent Nature hath hid in this Sublunary World, as it were the Instruments of Life and Soul, of Motion and Sense, of every thing" (Willis 1684, p. 3).

Sharpened with volatile salts of the blood until they are uniquely suited to their "offices of motion and sense" (Willis 1684, p. 15), they "stray," they "wander," cutting tracks through coarser brain matter due to their minute size and lightness. Indeed for Willis animal spirits are light itself, rarified to an even greater heat and delicacy than the vital flame that nurtures the body. Rather than being a predictable automaton, the body is so complex, so profusely animated by "self-moving energy," that Willis believes the ultimate *explanans* must be God. Nothing besides divine workmanship can be invoked to explain the perfection of organic design for its manifold purposes. However, like others in the Royal Society, Willis rejects the occultism of earlier treatments of spirit and seeks to redefine it in the terms of the new science. It is to this end he distinguishes between the two souls, one immaterial and immortal and the other earthly, the latter infused by God's grace throughout the matter of the body to create life.

The sensitive animal spirits, through the striations they leave on the soft matter of the brain, create the patterns of thought that constitute desire, memory and fantasy. It is their agency that drives the forms these patterns take, so that their emergent patterns are the sum of their distinct and individual actions. Animal spirits are both the medium and the message: activity is defined "as a blast of Wind in a Machine, being struck, [they] run hither and thither, so produce the Exercises of Sense and Motion in the whole body, or respective parts" (Willis 1684, p. 56.) Their work is diverse, for "the Animal Spirits love to expatiate themselves" (Willis 1684, p. 147); they meander through the callous body of the brain, preserving memories, through the brain's center, where they represent imaginings and fantasies, and into the streaked bodies, where they instigate their compatriots inside the nerves to action to generate acts of will. In death, "the Animal Spirits presently vanishing, after life is extinct, leave no Foot-steps of themselves" (Willis 1684, p. 24).

While Willis says repeatedly that animal spirits are the "immediate Instruments of thoughts," they are themselves under the control of the precordia, the system of organs regulated by the vital part of the corporeal soul, including the heart. The heart, in controlling the amount, viscosity, and velocity of the blood, has an important impact on the nature of the animal spirits – thus "Wisdom is much rather ascribed to the heart" (Willis 1684, p. 47). However, the precordia is under the thrall of the passions, which themselves influence the seat of the nervous system in the brain. The spirits are responsible for turning the feelings of the sensitive soul – tempered with the intellectual contributions of the rational soul – into action. But they are also responsible for gathering the sensory data from the external world that forms the basis for the passions, for the imagination, and for the images that are projected onto the inner recesses of the brain and on the basis of which the rational soul makes its choices.

3.4 Vital Medicine

Willis was one of a number of early modern physicians whose enthusiasm for iatrochemistry grew out of, and reinforced, his empirical method. "The business of a Physician and a Vintner," Willis wrote, "is almost the same: the blood and humors even as Wine, ought to be kept in equal temper and motion of Fermentation" (1684, p. 16). Indeed he claimed that he could "easily unfold the Curatory intentions, as also the effects and operations of every medicine, according to the Doctrine of Fermentation" (Willis 1684, p. 16). Following Paracelsus and van Helmont, Willis argues that chemical medicines have the power to alter the composition of the vital and animal spirits of the body, though he admits that how "the Animal Spirit is wrought in the Brain [is] very much in the dark" (1684, p. 14). He postulates that the part of the blood that is "highly volatile, spirituous, and endued with active Elements" (Willis 1681a, p. 87) is distilled in the blood vessels within the head until it is fine enough to enter the brain.

Earlier in the history of iatrochemistry Paracelsus relied heavily on analogies between the macrocosm and the microcosm, using heavenly events to explain earthly ones, and the fermentation processes of the natural world to explain those hidden in the body. Rejecting the Galenic system in which disease was caused by internal imbalance, iatrochemist physicians followed him in envisioning organic dysfunction as the result of the failure of internal *archei*, or seminal principles deep within the organs of the body, to catalyze the proper reactions (Debus 1998). Theorists based their descriptions of these processes on analogous processes in the natural world, since the reagents of the body could not easily be isolated nor their reactions observed. Acknowledging that in the case of the animal spirit it is mysterious "by what workman it is prepared, nor by what Channels it is carried, at a distance, quicker than the twinkling of an eye" (Willis 1684, p. 14), Willis relies heavily on macrocosm-microcosm analogies, and the modern reader must be careful in interpreting them.

The *Oxford English Dictionary* supplies a chronology of the term "analogy" that details its shift from "likeness" to the contemporary meaning, which more precisely refers to a comparison meant to highlight a similarity between two disparate quantities. Willis's frequent reliance on comparative conjunctions like "as it were" complicate attempts to interpret his iatrochemical descriptions as reducing one level of explanation to another, analogous, one. While Willis often compares the changing dynamics of the animal spirits to other varieties of fermentation which were explicable using the resources of iatrochemistry, he just as often uses the same syntactic structures to equate the actions of the spirits with the conscious behavior of human agents. In both cases he seems to be drawing a kind of likeness.

Rather than positing a reduction of animal spirits to corpuscles, he compares the symptomatology of the madman to the results of familiar alchemic processes in order to illuminate how, given the varied dispositions of the spirits, diverse "mad distempers" can arise. "The comparing of Animal Spirits with *Stygian* Water,"[15] for example, "clearly shews what is the conjunct or immediate cause of *Madness*"

[15] Or *Aqua fortis*, known today as nitric acid.

(Willis 1683, p. 202) – namely, the changing nature of the animal spirits from a gentle to an active and volatile disposition. In the case of melancholy, Willis offers an analogy between the chemical ecology of the melancholic brain and a vial of acid spirits of salt or vitriol.[16] The animal spirits react "like little acid atoms" (Willis 1683, p. 190) creeping about, with the sole aim of escaping the acetous liquid, until they have gathered outside its reach. Due to the conglomeration of animal spirits in certain areas of the brain, not only do certain thoughts become unavoidable for the melancholic, but they also take on an unrealistic significance and stature. The potency of the analogy is in its application of a known reaction – that of an acidic liquid and a salt or sulfur – to explain a class of symptoms as the result of an unknown reaction – that of the nervous juice and the animal spirits.

Yet when he comes to explain "the more remote or antecedent causes of *Madness*, *viz.* by which reason of which the Animal Spirits acquire a most sharp disposition," Willis cites how the spirits are "wont to be cast down by a violent and terrible passion" (1683, pp. 202–203). He expresses adamantly that animal spirits change their disposition "not by consent, nor from any force from another, but of themselves […]. To this vice of theirs, perhaps the Brain, or the Blood, or other parts may contribute somewhat, but the Spirits themselves are first and chiefly at fault" (Willis 1683, p. 201). Madness often comes about because the spirits give in to a powerful passion against the better judgment of the rational soul, or because they extend themselves too far out of hubris. The spirits of a madman are both "agitated like mad Bacchanals" (Willis 1683, p. 179) and "like to *Stygian*-Water," in that they are highly active and restless, "filled as it were with a *Nitrous Sulphur*, and indued with a notable mobility or unquietness" (Willis 1683, p. 201),[17] abandoning established tracks in the brain to make new ones with impunity. Rather than mechanical, the shifting character of the animal spirits is internally catalyzed, like a change in mood.

Instead of dismissing Willis's analogies between animal spirits and human agents as literary flair while heralding his analogies between animal spirits and chemical reagents as the birth of neuroscience, we should contextualize both sorts of "likenesses" within Willis's larger commitments. Spirit is the most active of the principles, and it imbues nature, Willis believes, with life force. The active powers of spirit are deeply explanatory, and Willis emphasizes that any mechanistic system that does not acknowledge them cannot capture the complexity of the natural world:

> "foreasmuch as it undertakes Mechanically the unfolding of things, and accommodates Nature with Working Tools, as it were in the hand of the Artificer, and without running to Occult Qualities, Sympathy, and other refuges of ignorance, and doth happily and very ingeniously disentangle some difficult Knots of the Sciences, and dark Riddles, certainly it

[16] Today's hydrochloric acid and sulfuric acid, respectively.

[17] For a discussion of the use of "as it were" and other indicators of simile in Willis and other early modern natural philosophers, see Harris (1917). Harris makes the compelling but potentially Whiggish argument that during this period analogy was used as a placeholder for a more technical explanation than contemporary knowledge could support. From this perspective, Willis's constant use of metaphor in the discussion of neuropathology would suggest that he did not find iatrochemistry sufficient to explain every aspect of the soul. For more on Willis and metaphor see Rousseau (2004, p. 4).

> deserves no light praise; but because it rather supposes, than demonstrates its Principles, and teaches of what Figure those Elements of Body may be, not what they have been, and also induces Notions extremely subtil, and remote from the sense, and which do not sufficiently Quadrate with the Phaenomena of Nature, when we descend to particulars, it pleases me to give my sentence for the third Opinion before-mentioned, which is of the Chymists" (Willis 1684, p. 2).

Willis rejects mechanism because the active nature of the principle of spirit, which finds its zenith in the animal spirits of the human body, is no metaphor.

While like many commentators M.J. Eadie dismisses Willis's animal spirits as "a fictitious entity in many ways analogous to the present day idea of the nerve impulse" (2003a, p. 14), he usefully divides Willis's categories of psychopathology into two classes: "disorders of intrinsically normal animal spirits" and "disorders of intrinsically abnormal animal spirits." Notably the more "mental" neurological dysfunctions, such as hysteria, frenzy, madness, and melancholy, are in the latter camp. In these cases the chemical nature of the animal spirits is altered – in melancholy, they grow dark and heavy, in madness sharp and strong – and so is their character. When bitten by a tarantula whose poison causes madness, Willis writes, victims convulse and writhe, displaying "unweariable dancing" (Willis 1681b, p. 46) and can only be cured by hearing music that soothes the convulsions into a more controlled movement. Willis explains the intervention as acting on the animal spirits, who "wander about hither and thither willingly" until, "delighted together" and enchanted by the music, they become exhausted from dancing and "at length rest from that madness" (1681b, p. 48). In this explanation, Willis has simply moved the symptoms with which the patient presents under the skin.

On a reductive interpretation of Willis's psychopathology, one would expect his interventions to be solely iatrochemical – melancholy could be countered by neutralizing the acidity of the blood, for example. But because Willis's neuropathology integrates the vitalization of the animal spirits with chemical etiology, he is able to maintain many of the cures traditionally targeted at the behavioral aspects of mental illness alongside more recent innovations drawn up on iatrochemical grounds. His first cure for melancholy is aimed at lifting up and "making volatile" the animal spirits through the withdrawal of the soul from strong passions and the participation in gently pleasant activities such as singing, dancing, and hunting. Alongside bloodletting and remedies aimed at returning the balance of the blood, Willis talks frankly about interventions aimed at cheering the spirits, who, it seems, retain an irreducible subjectivity that responds not just to chemical interventions but to humane ones. It is their disposition towards melancholy that is the target of therapeutic treatment.

3.5 Treating Spirits

In Willis's view, as demonstrated above, the disposition of individual spirits can have a powerful, even alarming, effect on their ability to, on the one hand, accurately transport a message, and, on the other, to navigate the brain. In this he was certainly not alone – as Sutton (1998, p. 33) has written, "Early modern spirits would be regularly personified as animalistic agents mischievous in the innards" – but in Willis's case the seeming anthromorphism played an explanatory role in his portrait of mental disorder. Malebranche, Descartes, and Locke all applied themselves to the challenge of using reason to overpower the spirits, and all agreed that were it not for the fall of man, will, imagination, and memory would be unnecessary. But Willis in particular emphasizes that psychopathology such as stupidity is *not* a fault of the rational soul, though it appears to be a breakdown of the intellect. Rather it is the fault of the animal spirits and their destruction of the imagination and the memory on which the intellect depends to maintain its rationality. The spirits of the sensitive soul, rather than the rational soul itself, go mad.

To cure their madness, Willis suggests dampening the symptoms of the disorder through correcting "the furies and exorbitances of the Animal Spirits" (Willis 1683, p. 206). Challenging the received Galenic picture he writes that it is "not so much in an adust bile or humor, or black and sharp vapour" that the causes of madness can be found, but rather in the degeneration of the animal spirits "from a gentle and benigne nature, as also a subtil and very active disposition, to wit, a *Spirituous-saline,* into another sharp, to wit, partaking of fluid *Salt*, and *Arsenical Sulphur*" (Willis 1683, p. 202). To take away the causes of the disorder, he suggests intervening on whatever is causing the sharpness of spirits in the first place through the use of medicines or remedies that alter the concentration of the active principles in the blood and thus adjust the processes of fermentation that may be compromising the spirits.

Yet the curative indications, which aim at relieving the symptoms of the disorder, are directed at the animal spirits as agents. Warnings, chidings, and punishments such as beatings "may suppress or cast down the Elation of the Corporeal Soul" (Willis 1683, p. 106). Thus phenomenological explanations are by no means excluded from his medical psychology but rather flow through it, integrated with the chemistry of the animal spirits, with the passions that drive the sensitive soul, and, more distally and opaquely, with the will of the rational soul.

I would submit that the relocation of phenomenal experience – the "what it is like to be"-ness – onto the spirits is not unique to Willis's psychopathology. Elsewhere in his writings desire and abhorrence follow the proclivities of the animal spirits, some of which are more "fugacious or apt to flight, or *pathetick*, or passionate" (Willis 1683, p. 184; emphasis in original). Sensions "Congruous and Curiously fitted to the Sensory" cause the animal spirits that receive them to rejoice, which in turn encourages the brain to "try several manifold endeavours" to achieve the object of desire (Willis 1683, p. 50). In a startling turn of phrase, Willis writes that the pursuit of desired objects is nothing but the animal spirits inviting the sensitive soul "to the worship of the Idol erected by themselves" (Willis 1683, p. 50). More directly their actions towards or away from their objects of desire bring changes to the corporeal body: drawing away from a perceived threat, the animal spirits pull the pores of the skin closed behind them, raising hairs. When pouring forth towards something beautiful, on the other hand, the spirits flood the precordia, causing a quickening of the blood and a feeling of expansiveness in the heart.

Much in the way that animal spirits are the true vessels for desire and repulsion, it is also they who are begging for nourishment when we are hungry and who are "the immediate Subject of sleep"; although the spirits of the cerebellum, who must keep the basic life functions consistent, do not get a "holy day" (Willis 1683, p. 87). We only wake when our spirits are rested, and our dreams are caused by the few spirits who escape the chains of sleep to wander "like Spectres in [the] Church-yard" of the imagination's tracks (Willis 1683, p. 94). Animal instinct and human habituation are a result of the fact that animal spirits are creatures of habit, in so far as it is easier for them to take a track already broken in than to cut their own path through the brain. Spirits of the insane, "being very fierce and provoked, both fortify the imagination [....] and actuate also the *Pracordia* [*sic*] with great vigor" (Willis 1683, p. 205) – such that the madman himself becomes fierce and provoked. The case of the drunkard is similar. When the animal spirits become melancholy, they move as little as possible and stick to the least demanding paths, leading to a narrowing of thought.

Willis gave a set of lectures at Oxford in 1663 that were carefully transcribed into a notebook by the student John Locke. Within pages recording Willis's reflection that the brain of a newborn is a *tabula rasa*, Locke records a lecture on pleasure and pain. "Pain," Willis declares, "is an annoying sensation just as pleasure is an agreeable one" (1980, p. 67). But what is an agreeable sensation? It is, Willis argues, the "gentle," "pleasing," and "agreeable" dances of the spirits – a pleasant motion. What would today be called the quale of pleasure is not explained away, or reduced to a particular sort of action – fast or slow, hard or soft. It is the "agreeable motion" of the spirits themselves that "recall past pleasures through the memory" (Willis 1980, p. 67) in reflexive delight. The blissful state of the spirits causes the gross motions of the body associated with pleasure, such as smiling and salivation. What causes the particular type of motion considered "pleasant"? What property links the stimulus to the resulting actions of the spirits? What does it mean to say that the spirits have preferences, tastes? Willis does not address this other, harder, problem.

When Willis, in the following paragraph, describes pain as arising from the dissipation of the animal spirits due to an external object that causes the dissolution of a continuum in the body, the disordering of the motion of the spirits due to spasms or another cause, or the defect of an organ, the identification of pain with pleasure makes it clear that the dissipation of the spirits can only be described as "an unpleasant motion" (Willis 1980, p. 67). Their dissipation is what causes pain to the body that houses them – it is not what pain is.

3.6 Neuropathology's Mad Core

I have argued against the view that Willis's animal spirits, while a symbolic nod to an antiquated notion of the soul, in fact shifted the paradigm of medical psychology towards reductionism by anticipating today's neural synapses. My case has been made on the grounds that the phenomenology of the spirits has causal efficacy in Willis's descriptions of mental illness, as well as in other physiological functions.

One might respond that despite his personification of animal spirits, Willis must be read as a reductionist because his explanations of mental illness are often presented in iatrochemical terms. Here caution is needed, however, as argued above: Willis frequently uses analogies from chemistry to describe "like" processes that occur within vital reactions in the body, as well as in the conscious mind of an agent.

Iatrochemical explanations did not exclude phenomenological ones for Willis. Rather the situation of the animal spirits within the sensitive soul allowed him to describe the lived experience of mental illness in causal terms, by rendering it compatible with lower-level explanations on the corpuscular level. Nowhere in his corpus does he show an ambition to replace his descriptions of the psychopathological dispositions of animal spirits with purely chemical ones. Rather, iatrochemistry is introduced to explain transformations within the sensitive soul to which the spirits are subject *on account of* their own natures. To the contemporary reader Willis's spirits can seem like a diverting and whimsical metaphor for the menagerie of neurotransmitters, pheromones, and other messengers inhabiting the modern brain. As has been demonstrated, however, for Willis it was more often the iatrochemical explanations that were analogical, borrowing observations from visible world to aid with hypotheses about the secret inner world of the spirits. While a contemporary neuroscientist might put more faith in the ontology of the brain's chemistry, Willis found the phenomenology of neuropathology to be a proper object of medical investigation and explanation.

Evidently one must look elsewhere in the prehistory of psychiatry to find the first attempts to exclude entirely the phenomenological aspect of madness, the concern for the lived experience of the loss of reason. It is interesting to note, however, that anxieties about the cost of such a reduction were present amongst early modern critics. Henry More, for example, questioned how animal spirits could be capable of anything beyond mechanical motion: he found them "utterly incapable of *Memory*… it is impossible to conceive *Memory* competible to such a subject, as it is how to write Characters in the water or the wind" (quoted in Sutton 1998, p. 145). More's critique was directed against the mechanistic spirits of Descartes, and he would perhaps have been more amenable to the vitalistic ones of Willis. Likewise for the modern critic of biomedical psychiatry, Willis's strange preservation of the "human" aspect of mental illness in the form of our animal spirit familiars may suggest subtle challenges to the contemporary mechanistic turn in psychiatry, ushered in with cognitive neuroscience and genetics. With all its nefarious and subtle allusions to the bestial, the angelic and the alchemical, the term "animal spirit" has proved, in Sutton's words, "awkward and productive" (1998, p. 32) – with it Willis founded a new science even while recognizing the indelible subjectivity of the mad soul.

References

Aristotle (1984) In: Barnes J (ed) The complete works of Aristotle: vol I. Princeton University Press, Princeton

Canguilhem G (1977) La formation du concept de réflexe aux XVII^{e} et XVIII^{e} siècles. J. Vrin, Paris

Clericuzio A (1994) The internal laboratory: the chemical reinterpretation of medical spirits in England (1650–1680). In: Rattansi P, Clericuzio A (eds) Alchemy and chemistry in the 16th and 17th centuries. Kluwer Academic, Dordrecht/Boston, pp 51–84

Conry Y (1978) Thomas Willis ou le premier discours rationaliste en pathologie mentale. Revue d'histoire des sciences 31:193–231

Debus AG (1998) Chemists, physicians, and changing perspectives on the scientific revolution. Isis 89:61–81

Descartes R (1650) The passions of the soul. J. Martin and J. Ridley, London

Descartes R (2003) Treatise of man (trans: Hall TS). Prometheus Books, Amherst

Eadie MJ (2003a) A pathology of the animal spirits – the clinical neurology of Thomas Willis (1621–1675): Part I – Background, and disorders of intrinsically normal animal spirits. J Clin Neurosci 10:14–29

Eadie MJ (2003b) A pathology of the animal spirits – the clinical neurology of Thomas Willis (1621–1675): Part II – Disorders of intrinsically abnormal animal spirits. J Clin Neurosci 10:145–157

Frank RG (1990) Thomas Willis and his circle: brain and mind in seventeenth-century medicine. In: Rousseau GS (ed) The languages of psyche: mind and body in enlightenment thought. University of California Press, Berkeley, pp 107–146

Harris F (1917) From the vague to the concrete in science and medicine. Can Med Assoc J 7:865–878

Hunter RA, Macalpine I (1963) Three hundred years of psychiatry: 1535–1860. Oxford University Press, Oxford

Jarcho S (1982) Blankaart's dictionary: an index to 17th-century medicine. Bull N Y Acad Med 58:568–577

Kassler J (1998) Restraining the passions: hydropneumatics and hierarchy in the philosophy of Thomas Willis. In: Graukroger S (ed) The soft underbelly of reason: the passions in the seventeenth century. Routledge, London/New York, pp 147–164

King LS (1978) The philosophy of medicine: the early eighteenth century. Harvard University Press, Cambridge, MA

Meyer A, Hierons R (1968) On Thomas Willis's concepts of neurophysiology, part I. Med Hist 9:1–15

Ochs S (2004) A history of nerve functions: from animal spirits to molecular mechanisms. Cambridge University Press, Cambridge

Rather LJ (1974) Pathology at mid-century: a reassessment of Thomas Willis and Thomas Sydenham. In: Debus AG (ed) Medicine in seventeenth-century England: a symposium held at UCLA in honor of C. D. O'Malley. University of California Press, Berkeley, pp 71–112

Rocca J (2007) William Cullen (1710–1790) and Robert Whytt (1714–1766) on the nervous system. In: Whitaker H, Smith CUM, Finger S (eds) Brain, mind, and medicine: essays in eighteenth-century neuroscience. Springer, New York, pp 85–96

Rousseau GS (2004) Nervous acts: essays on literature, culture and sensibility. Palgrave Macmillan, Houndmills/Basingstoke/Hampshire/New York

Sutton J (1998) Philosophy and memory traces: Descartes to connectionism. Cambridge University Press, Cambridge

Thomson A (2008) Bodies of thought: science, religion, and the soul in the early enlightenment. Oxford University Press, Oxford

Wallace W (2003) The vibrating nerve impulse in Newton, Willis and Gassendi: first steps in a mechanical theory of communication. Brain Cogn 51:66–94

Williams AN (2002) "Of stupidity or folly": Thomas Willis's perspective on mental retardation. Arch Dis Child 87:555–558

Willis T (1681a) The remaining medical works of that famous and renowned physician Dr. Thomas Willis. Thomas Dring, London

Willis T (1681b) An essay of the pathology of the brain and nervous stock: in which convulsive diseases are treated of. Thomas Dring, London

Willis T (1683) Two discourses concerning the soul of brutes which is that of the vital and sensitive of man: Englished by S. Pordage. Thomas Dring, London
Willis T (1684) A medical-philosophical discourse of fermentation, or, of the intestine motion of the particles in every body. Thomas Dring, London
Willis T (1980) Thomas Willis's Oxford lectures. Dewhurst K (ed). Sanford Publications, Oxford
Wright JP (1991) Locke, Willis, and the seventeenth-century Epicurean soul. In: Osler MJ (ed) Atoms, pneuma and tranquillity: epicurean and stoic themes in European thought. Cambridge University Press, Cambridge, pp 239–258
Zilboorg G, Henry GW (1941) A history of medical psychology. Norton, New York

Chapter 4
Hooke's Mechanical Mind

J.J. MacIntosh

4.1 Background

Seventeenth century thinkers inherited a doctrine of perception, and hence a doctrine of thought, from their mediaeval precursors. The mediaevals had inherited from Plato the view, which they ultimately dismissed in favour of an Aristotelian alternative, that vision required an emanation from the eye which made contact with the object and returned the information such contact gave rise to the brain.[1] But there were all too obvious problems with this "visible breath," as Adelard of Bath (c. 1080–1152) called it. How, for example, did the ray know when it had reached the appropriate object, and how did it know to whom to return? And how, in the case of (say) heavenly bodies did it get so far, so fast? And why didn't the rays from your eyes and those from mine become entangled? The most important argument, however, was that of Alhazen,[2] who pointed out that there was a principle of parsimony involved: since seeing requires light, and since it assumes *something* coming to the percipient from the perceived, we can assume such incoming information is available without its having to be excited by an emanation from the eye. As John Pecham (1230–1292) said, explicitly "following in the footsteps of" Alhazen, "it is superfluous to posit such rays."[3] By the mid-seventeenth century this was the received view. "Sight is made by Reception, and not by Extramission," said Sir Thomas Browne.[4] Nonetheless, Plato's "extramission" view of vision, offered in more detail by Euclid

[1] Plato *Timaeus*, 45b–46a, and cf. *Theaetetus* 156d–e.

[2] Abu Ali al-Hasan ibn al-Haytham (Alhazen), c. 965–1038/9.

[3] Pecham 1970, p. 127.

[4] Browne 1646, iii. vii. p. 120. Browne's position is not completely clear, since he was talking about the Basilisk, but he appears to feel that his remarks can be generalized, and indeed if the claim applies to the Basilisk it can scarcely fail to apply to less harmful optical interaction.

J.J. MacIntosh (✉)
Department of Philosophy, University of Calgary, 2500 University Drive, NW,
Calgary, AB T2N 1N4, Canada
e-mail: macintos@ucalgary.ca

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_4,

and Ptolemy,[5] continued to have an effect, and may be seen to influence Hooke's presentation of mind body interaction.

Leaving aside, for the moment, the question of the *mechanism* of information transmission, clearly information somehow reached the percipient. This information was thought to come in the form of *images*,[6] which were then processed in the brain. This processing involved two clear tasks. First, somehow, the processing system had to *distinguish* the input, to notice that milk, for example, was both sweet *and* white, and also, it had to *collect* or *conjoin* the input, to notice that the information coming from the various senses was information from a *single* entity. Moreover, some subsystem had to allow us to judge that it is the sense of sight that sees, of hearing that hears, etc., for this is not something that can be sensed.[7] For these processing tasks a theoretical entity, the *common* sense, was invoked: an internal sense that had something in common with all the external senses, and so could process the information they severally provided. Roger Bacon located this *sensus communis* in the first of the brain's three "chambers, cells, parts [or] divisions."[8] This became the traditional division, with the second ventricle being devoted to thought and judgment, and the third to memory.[9]

This information, once processed, must be retained: "there [is] another faculty of the soul…, the function of which is to retain the species coming from the particular senses, … which is called imagination and is the coffer and repository of the common sense … the whole faculty … composed of these two … is called phantasia or the *virtus phantastica*. … [S]ince the common sense receives the species, and the imagination retains it, a complete judgement follows regarding the thing, a judgement formed by phantasia."[10]

Following Aristotle, the mediaevals also held that an *immaterial* faculty was required given the ability of humans to *abstract* from particular events. Thus, looking at a pen, say, we not only have a visual impression of a cylindrical coloured object, we see it *as* a *cylinder*, and indeed, *as* a *pen*. Such an ability to perceive the immediate input as *categorized* was referred to yet another theoretical entity, the agent intellect.

---

[5] See Lindberg 1976 for details. Lindberg suggests (ch. 3) that there were three main considerations involved, mathematical, physical, and physiological, which led to various versions of three main outlooks on vision, not integrated until the work of Alhazen.

[6] The transmitted packets of information received a variety of names, of which 'form,' 'species,' 'phantasm,' 'image,' and 'idea' were perhaps the most common in the early modern period. Earlier, Roger Bacon (1214–1292/4) offered the following as synonyms: '*lumen*', '*idolum*', '*phantasma*,' '*simulacrum*,' '*forma*,' '*intentio*,' '*similitudo*,' '*umbra,*' '*virtus*,' '*impressio*,' and '*passio*' (see Lindberg 1976, p. 114), and see Crombie 1967, p. 5, n7 for further examples.

[7] We don't see, hear, etc. that we are perceiving by seeing, or hearing, or touching, etc.

[8] Bacon 1900, Pt. V, Optic, part 1, Dist. 1, ch. 2; Il, 5; also available in part in Grant 1974. I have used Lindberg's translations in Grant 1974.

[9] Crombie 1967, p. 69. For further details see "Beginnings: ventricular neuropsychology," by C.U.M. Smith, Chap. 1 of the present work.

[10] Bacon, op cit.

4.2 Descartes[11]

These views were taken over in effect by Descartes, though with some significant alterations.[12] The mediaevals were puzzled by the fact that the retinal image is – compared with what appears to be our perception of the world – inverted. Kepler was apparently the first to see that this was irrelevant, since the information transmitted from the eye to the brain via the animal spirits was not transmitted as an *optical* image:

> the thin hollow nerve is not optically straight; and if it were, it would nevertheless immediately become crooked because of the twisting of the eye, and its opaque parts would become opposed to the tiny opening or entrance into the passage way. Therefore light neither passes through, nor is refracted at, the posterior surface of the vitreous humour, but impinges there.[13]

Descartes retained the terminology of images, but took Kepler's point about "the nature of these images," which, he said, must be conceived in a manner quite unlike that in which "our philosophers" conceive them.[14] Moreover, Descartes, as a corpuscularian,[15] was happy to allow that the transmission was purely mechanical,

---

[11] For important, detailed, and interesting accounts of Descartes's views see Clarke 2003, and Brown 2006.

[12] See MacIntosh 1983 for further details.

[13] *Paralipomena*, Kepler 1604, V.2, 169. The point was still not clear to all the members of the Royal Society three quarters of a century later:

> The minutes of July 31 [1679] were read: whereupon there was a further discourse about shortsighted persons, and of the ways of vision, from the assertion of Mr. HOOKE, that a man used to see things always inverted would in time judge, that he saw them as they are. Dr. CROONE queried, whence it should come, that the conception should imagine that object erect, which is represented at the bottom of the eye inverted. Dr. GREW supposed that it might proceed from the   of the optic nerves, which might cause a second inversion. Mr. HOOKE thought that this could not be the cause, since it was not general in all creatures, and he conceived, that the inversion of the optic nerve was in none observable: but that it rather proceeded from the mind's making comparison of the sensation by the eye with the sensation made by the touch: or rather, that it is an idea or the rule of sight implanted in the soul by nature. (Gunther 1930, 7:530; the blank space is Gunther's.)

[14] *Dioptrique* 4, Adam and Tannery 1964–1976, 6.112, Olscamp 1965, 89.

[15] Robert Boyle's term, which allowed him to gather plenists such as Descartes and vacuists, including atomists, under a common anti-scholastic banner as thinkers who were willing to ground all natural phenomena in the interaction of minute particles of matter. In the *Origine of Forms and Qualities*, Boyle mentions "that Philosophy, which, I find, I have been much imitated in calling *Corpuscularian* (Boyle 1999–2000, 5:289)." For Boyle's justification of the umbrella term see *Certain Physiological Essays*, Boyle 1999–2000, 2:87. Leibniz, writing in 1669, agreed that the term captured a wide variety of thinkers:

> At the beginning I readily admitted that we must agree with those contemporary philosophers who have revived Democritus and Epicurus and whom Robert Boyle aptly calls corpuscular philosophers, such as Galileo, Bacon, Gassendi, Descartes, Hobbes, and Digby, that, in explaining corporeal phenomena, we must not unnecessarily resort to God or to any other incorporeal thing, form or quality (*Nec Deus intersit, nisi dignus vindice nodus inciderit*), but that, so far as can be done, everything should be derived from the nature of body and its primary qualities—magnitude, figure, and motion." ("The Confession of Nature against Atheists," Gerhardt 1875, 4:106; Loemker 1969, p. 619.)

though he retained, initially, the notion that light is transmitted by pressure rather than emission, offering us the analogy of a blind person's stick which transmits information but not matter.[16]

Here then we have what by the second half of the seventeenth century had become the standard account:

(i) Information comes from external objects to the percipient, not in the form of Aristotelian images, but as corpuscles (or 'pressure') impinging on the various nerve endings.
(ii) The information was then carried to the brain by the animal spirits, which were, in effect, theoretical entities assumed to travel via the nerves to the brain.[17]
(iii) Arriving in the brain they affected the brain to produce ingrained memories (Descartes's *plis de mémoire*, memory creases), which were stored in the brain; however information was also transmitted to the incorporeal mind via the conarion, or pineal gland, and free willed decisions to move were, in humans, similarly transmitted from the mind to the brain, and thence to the various muscles involved.

4.3 Standard Difficulties with the Cartesian Model

There were standard difficulties with the Cartesian model. I mention three:

(i) interaction, and the implausibility of the conarion (pineal gland) model;
(ii) the implausibility of Descartes's quantity of motion solution; and
(iii) the problem posed by event memory.

Though remaining a substance dualist,[18] Robert Boyle nonetheless saw clearly the problems involved in the first two areas. For, said Boyle, "this Union of an

[16] Descartes begins *La Dioptrique* with the suggestion that light is transmitted by pressure, and offers, in a letter to Mersenne a further analogy: this transmission is analogous to the 'pressure' exerted by a whirling stone in a sling on the cord to which it is attached. ("la lumière … presse la matiere subtile en ligne droite ver nos yeux … qu'une pierre qui est tournée en rond dans une fronde, presse le milieu de cete fronde, & tire la chorde en ligne droite par le seule force de son mouuement circulaire.") (Descartes to Mersenne, August 27, 1639, Adam and Tannery 1964–1976, 2:572.) However, as J. F. Scott points out (Scott 1952, p. 33), his explanations of the different properties of light, and particularly colour, seem already, in Discourse I, clearly to require an emission theory of light.

[17] Robert Boyle remarked of these "minute" and "invisible" spirits that "prying Anatomists have not been able in dissected Nerves to discern so much as the channels, through which they pass; yet those Invisible Spirits, conveyed (or impelled) from the Brain to the Nerves, serve to move in various manners the Lims, and even the unwieldy bodies themselves of the greatest Animals, and to carry them on in a progressive motion for many hours together, and perhaps enable them to spring into the Air, and move through it by leaping; though divers of these Animals weigh many hundred, and others several thousand of pounds." *Languid and Unheeded Motion*, Boyle 1999–2000, 10:265. See further Clarke 1968.

[18] So I say, because he assumes the immateriality of the mind, and speaks in what amounts to a dualistic way about the mind. Still, in 1666, he did add to his criticism of the notion of substantial forms: "when ever I shall speake indefinitely of Substantiall forms, I would alwayes be understood to except

Incorporeal with a Corporeal Substance … is a thing so unexampled in Nature, and so difficult to comprehend, that I somewhat question, whether the profound Secrets of Theology, not to say the adorable Mystery it self of the Incarnation, be more abstruse than this." And, he asks, how can "any part of the Body, without excepting the Animal Spirits, or the *Conarion* … make Impressions upon a Substance perfectly Incorporeal"? Nor will it do to say, as Descartes suggested, that it only the *direction* and not the *quantity* of motion that is involved, for both are mysterious: "Nor is it a small difficulty to a meer Naturalist (who, as such, does not in Physical matters take notice of Revelations about Angels,) to conceive, how a finite Spirit can either move, or, which is much the same thing, regulate and determine the motion of a Body." Even if all these problems were solved, there is still the problem involving what we, though not Boyle, refer to as qualia: why do some impressions on the conarion produce visual sensations and others auditory sensations? And why do they produce just the precise sensations they do? "What can be answered, but that it was the good pleasure of the Author of Humane Nature to have it so?"[19]

Finally, there is the third issue: how is event memory to be accommodated? Descartes distinguished event memory, which like his mediaeval predecessors he took to be a neurophysiological phenomenon, from *intellectual* memory, which required abstraction, and which was a function of the incorporeal mind.[20] This led Constantijn Huygens to write to Descartes wondering whether (and if so, how) he could remember his friends and his earlier life after death. Descartes's reply was clearly meant to be comforting: "Those who die pass to a sweeter and more tranquil life than ours; I cannot imagine otherwise. We shall go to find them someday, and we shall still remember the past, because we have, on my view, an intellectual memory which is certainly independent of the body."[21] There is, however a problem: elsewhere Descartes makes it clear that intellectual memory is *not* memory of individual events: "this intellectual memory has universals rather than particulars as its objects and so cannot enable us to recall every single thing we have done."[22]

---

the Reasonable Soule, that is said to inform the humane Body; which Declaration I here desire may be taken notice of, once for all (*Forms and Qualities*, Boyle 1999–2000, 5:300)." Boyle seems to have been unaware of the work of Steno and Swammerdam which made Descartes's choice of the pineal gland seem empirically implausible. (For a discussion of Steno's work, see Olden-Jørgensen 2009, pp. 153–154.) Swammerdam is not mentioned at all in Boyle's published works. Steno is apparently referred to obliquely by Boyle in *Final Causes* (Boyle 1999–2000, 11:105, and directly along with Malpighi in the *Porosity of Bodies* (Boyle 1999–2000, 10:110). However, *Porosity of Bodies* was published in 1684, 5 years before Malpighi's 1689 *De Structura glandularum conglobatarum*, and Boyle says merely that "the excellent Anatomists *Steno and Malpighi* are said to have discovered" "numerous glandules … called *Glandulae miliares*," which seems to suggest something other than a first hand acquaintance with Steno's work. In geological matters, Yamada 2009 (pp. 116–117) suggests that, possibly as a result of Steno's mentor Ole Borch's meetings with Boyle in 1663, knowledge of Boyle's views may have had an influence on Steno's geological work.

[19] *Excellency of Theology*, Boyle 1999–2000, 8:68.

[20] "Besides this [event] memory, which depends on the body, I distinguish (*reconnais*) another one, entirely intellectual, which depends on the soul alone." (Descartes to Mersenne, April 1, 1640, Adam and Tannery 1964–1976, 3:34)

[21] Descartes to Huygens, Oct 10, 1642, Adam and Tannery 1964–1976, 3.580.

[22] Adam and Tannery 1964–1976, 5.150, Cottingham 1976, p. 9. Descartes in general glosses over the problem of memory. When Burman, for example, suggests that, if we are being sceptical

There are really two difficulties here. How can intellectual memory, which is of universals not particulars, enable us to remember particulars and, additionally, how are such memories stored in an incorporeal mind? Descartes was aware of the second difficulty. In 1644 he wrote: "As for memory, I think that the memory of material things depends on the traces which remain in the brain, after any image has been imprinted on it; and that the memory of intellectual things depends on some other traces which remain in thought itself. But the latter are of a wholly different kind from the former, and I cannot explain them by any illustration drawn from corporeal things without a great deal of qualification."[23] Unfortunately, no doubt, but also unsurprisingly, Descartes never did get round to solving either difficulty.

4.4 Descartes's Dualistic Legacy

I shall suggest that, probably at least, Hooke was a (closet) materialist. There were, however, a number of social and other pressures to prevent Hooke from straightforwardly *saying* that he thought the mind was material.

In the early modern period a variety of reasons were offered for the immateriality of the soul; here I consider two main ones, one intellectual and one spiritual, plus a third, personal, one for Hooke. The intellectual one involved the need for an immaterial soul to account for certain intellectual abilities such as our ability to abstract, a point already offered in their different ways by Plato and Aristotle. Another, continuing, favourite was our ability to comprehend various mathematical results which could not be reproduced in matter, such as the incommensurability of the sides and diagonal of the square. Hooke's patron, Robert Boyle, interested in, not to say obsessed by, the incommensurability result, found a variety of mathematical results interesting for much the same reason. Boyle mentions "the consequences of the incommensurableness of the side and Diagonall of a square, as also of the sixteenth proposition of Euclids Third Book[24]; on which occasion one may perceive an

about our knowledge, we should notice the possibility that "the weakness of memory" may lead us astray, Descartes replies, "I have nothing to say on the subject of memory. Everyone should test himself to see whether he is good at remembering. If he has any doubts, then he should make use of written notes and so forth to help him." (Adam and Tannery 1964–1976, 5:148, Cottingham 1976, p. 5; see further Cottingham's comments at pp. xxvii–xxviii.)

[23] Descartes to Mesland, May 2, 1644, Adam and Tannery 1964–1976, 4:114.

[24] *Elements* III.16 deals with the angle between the tangent and arc of a circle, which puzzled classical, mediaeval, and early modern thinkers. It seems visually clear that the angle of contact between a tangent line and the arc of its circle has a magnitude. But given two magnitudes, if you reduce the larger by more than half and continue this process with the residue you will eventually arrive at a magnitude less than the originally smaller magnitude (Archimedes' axiom). Now suppose a straight line intersecting the tangent at the point of contact. That line will cut the circle at a further point and thus make an angle with the tangent line larger than the angle of contact. This will continue to remain true after any halving, giving rise to an apparent paradox. Boyle refers to this point also in *Reason and Revelation*, Boyle 1999–2000, 8:266. See further Heath 1956, 2:39–43, and Jesseph 1999, pp. 159–173. The problem was dissolved by Möbius in 1846 (see Hofmann 1974, pp. 12–13 and 13 n8, and Möbius, 1885, 2:4–5).

operation of the mind which seems to transcend matter, what texture soever we suppose it to be put into."[25] The existence of asymptotes provided a further case.[26]

In the early eighteenth century Leibniz offered a thought experiment to show the foolishness of the notion of matter thinking: suppose, he wrote, "there were a machine, so constructed as to think, feel, and have perception, it might be conceived as increased in size, while keeping the same proportions, so that one might go into it as into a mill. That being so, we should, on examining its interior, find only parts which work one upon another, and never anything by which to explain a perception."[27]

The point was not generally thought doubtful. Earlier Ralph Cudworth and Richard Bentley thought the matter to be clear: "Omnipotence it self cannot create cogitative Body," said Richard Bentley in his Boyle lectures of 1691, and Cudworth had already made the same claim: "It is demonstrably evident and mathematically certain, that no Cogitation can possibly arise out of the Power of Matter."[28]

Locke, however, as is well known, was an exception to the standard belief that matter could not, logically could not, think. Locke held that, as far as he could see, God could superadd thought to matter if he chose.[29] Locke was simply mystified by our ability to think, and felt that neither corporeality nor incorporeality were much help in this conceptual thicket. "Pray tell us," he wrote in the margin of Burnet's *Third Remarks*,[30] "how y^u^ conceive cogitation in an unsolid created substance. It is as hard, I confess, to me to be conceived in an unsolid as in a solid substance."[31]

And Locke was not the only exception. The French scientist J B Duhamel wrote to Oldenburg in 1673:

> I see that [non-human animals] have feeling and consciousness, although the manner in which they are conscious is unknown to me … . Thus, although there is nothing more shocking … than granting some consciousness to matter, it nevertheless seems to me that we should first … see whether animals who have every appearance of being nothing but matter do not have some consciousness, rather than be dismayed by the absurdities, either real or apparent, which occur in this matter.[32]

---

[25] Boyle 2006, 3.5.9 (BP 1990, 7:162v), See also, e.g., *Christian Virtuoso I*, Boyle 1999–2000, 11.297, 337–38. Aristotle offers a further reason for such an interest: "the pain from thirst is opposed to the pleasure from drinking, but there is none opposite to the pleasure from contemplating the incommensurability of the side and the diagonal (*Topics* 106a36ff.)." A pleasure, in short, with no concomitant pain.

[26] See, e.g., *Things Above Reason*, Boyle 1999–2000, 9:373: where Boyle mentions "the Angle of Contact, the Doctrine of *Asymptotes*, and that of surd numbers and incommensurable Lines, all which … perplex the greatest Mathematicians."

[27] Leibniz, *Monadology*, §17, Gerhardt 1875, 6:609.

[28] Bentley 1699, p. 62. Cudworth 1731, p.302.

[29] Locke 1975, *Essay* 4.3.6.

[30] Burnet 1699.

[31] Porter 1984, p. 48; also Burnet 1989, p. 85. Leibniz explicitly criticized Locke's view in his posthumously published *Nouveaux Essais*, 4.3.6. Leibniz 1962.

[32] Jean Baptiste Duhamel to Oldenburg, Oct. 16/26, 1673, in Oldenburg 1965, 10:298, 299–300.

and Pierre Bayle noted that "It is a pity that Descartes' view is so implausible and difficult to sustain, because it is otherwise very advantageous to the true faith, which is the only reason why some people continue to hold it."[33]

Bayle's remark brings us to the second main reason why the soul was held to be immaterial. In the early modern period the soul's immateriality had two interlocking functions: first, its immateriality allowed at least the *possibility* of immortality (subject always to God's pleasure). In *The Christian Virtuoso, Part I*, Robert Boyle remarks, "After the Existence of the Deity, the next grand Principle of Natural Religion, is, the *Immortality of the Rational Soul*; whose genuine consequence is, the Belief and Expectation of a Future and Everlasting State."[34] Secondly, it separated us from creatures lacking free will: "Those who deny that there is any thing Incorporeal in man, make him but a mere Engine and little better than a Wind-mill, whose Essential Frame <is> but a Mechanical Contrivance; and all whose Motions or Functions, depend upon the Impulse of an External Body, the Wind."[35] Spinoza suggested that the distinction between free and unfree actions was that the causal principle was internal in the case of free actions, external in the case of unfree actions, and a formally similar point is found in Leibniz, but Kant was unimpressed: this is the freedom of the turnspit which, once it is wound up, is free to wind down.[36]

There was also, for Hooke, a third, perhaps equally important, reason for publicly accepting the soul's immateriality. He saw himself as being straightforwardly an adjunct of Robert Boyle, who was dedicatedly pious.[37] Hooke at one time worked directly for Boyle, and even after Boyle found him regular employment with the Royal Society, continued to dine regularly with Boyle and Boyle's sister Katherine.

So Hooke had personal as well as societal reasons for accepting, or apparently accepting, the standard view. However, I confess I wonder whether he really did.

4.5 Hooke's End Run Around the Problem

Hooke showed himself willing to ignore the all too apparent problem concerning the interaction between the immaterial mind and the material brain. What was of interest, Hooke clearly thought, was what we could reasonably say about the neurological situation. Here he shows himself willing to argue by analogy and, applying various mechanical possibilities rather freely, willing to produce

[33] "C'est dommage que le sentiment de M. Des Cartes soit si difficile à soutenir, & si éloigné de la vraisemblance; car il est d'ailleurs très-avantageux à la vraie foi, & c'est l'unique raison qui empêche quelques personnes de s'en départir." Article "Rorarius," Bayle 1734, 4:906.

[34] *Christian Virtuoso I*, Boyle 1999–2000, 11:297.

[35] Boyle 2006, p. 249, §3.5.4 (BP 1990, 1:36).

[36] Kant 1993, p. 101.

[37] For the interaction between Boyle's scientific and religious views see MacIntosh 1992.

a model, or quasi-model which, if not showing precisely *how* the brain actually worked, would at least show enough for us to see that it *could* have a mechanical explanation.[38] Like his patron, Boyle, Hooke seems to have been (comparatively) indifferent concerning the *realism* of his model, in this case at least.[39] He offers an account of how the soul can operate, but it is an account filled with analogies that are meant simply to soothe the reader into following Hooke's account a little further. Moreover, it is a straightforwardly material model:

> Now because nothing is so well understood or apprehended, as when it is represented under some sensible Form, I would, to make my Notion the more conceivable, make a mechanical and sensible Figure and Picture thereof, and from that shew how I conceive all the Actions and Operations of the Soul as Apprehending, Remembring and Reasoning are performed.[40]

Here are the main points:

1. "[T]here may be a certain Place or Point somewhere in the Brain of a Man, where the Soul may have its principal and chief Seat."[41]
2. It is to this part of the brain that "all the Impressions made from the Senses upon adapted Matter [are] delivered."[42]
3. These "Impressions … are no other but actual Locomotions given to the Parts of Matter or Bodies so or so moved."[43]
4. We may assume that there are various kinds of matter in the brain which are "adapted to receive the Impressions from the five Senses," that is, a kind of matter for receiving and storing visual impressions, another for auditory impressions, and so on.
5. Hooke offers a pair of experimental results to show that this supposition is plausible. We know, for example, that light may be stored in inorganic matter, as shown by the Bolonian Stone, and by Balduin's compound of "Chalk and Niter."[44] Since we see this to be possible in general, "it may yet possibly be done much more powerfully and effectually by the Chymistry of Nature in the

[38] Or so I call it, because that is how Hooke describes it. However, there are some difficulties with the notion that it is *strictly* mechanical since, as we shall see, it depends on an analogy with Hooke's notion of light, which is not in any straightforward way mechanical. For an important and detailed discussion of Hooke and mechanism see Henry 1989.

[39] See, e.g., Boyle, *Spring of the Air*, Boyle 1999–2000, 1:165–6. For Hooke see, besides the present case, his vibratory model for gravity in *Of Comets and Gravity*, Hooke 1705, pp. 176–86.

[40] Hooke 1705, 141.

[41] Hooke 1705, 141.

[42] Hooke 1705, 141.

[43] Hooke 1705, 141.

[44] Calcium nitrate. The Bolonian Stone, the first known case of inorganic phosphorescence, was prepared from crumbled and heated barite (barium sulphate) in 1602 by Vincenzo Casciarolo, a cobbler who dabbled in alchemy. When exposed to light it acquires the ability subsequently to glow in the dark, having "absorbed the golden light of the sun, like a new Prometheus stealing a Celestial Treasure (Licetus 1640, quoted Roda 1999)." See further Harvey 1957 and Roda 1999. For Boyle on Kraft, Balduin and the phenomenon of phosphorescence see *The Aerial Noctiluca* (Boyle 1999–2000, 9:265– 341), and for Boyle's account of a visit from Kraft see Boyle 1999–2000, 9:441–452. Boyle's

Digestions and Preparations made in the wonderful Elaboratory of the Animal Body."[45] Similarly in the auditory case, we are familiar with "the Unison-toned Strings, Bells or Glasses, which receive Impressions from Sounds without, and retain the Impression for some time." These examples "which I am fain to bring for Explication only," suggest that there can be matter adapted to receive and retain impressions from *all* our senses. So the material storage of sensory information can be assumed as a clear possibility.[46]

6. Surrounding this part of the brain ("which I will henceforward call the Center") is a sphere of matter adapted for forming, receiving, and storing ideas. These ideas are "material and bulky … Bodies of determinate bigness … impregnated with determinate Motions." Since they are material, they occupy space in the brain and have distinct qualities, shapes, and motions. These ideas are formed by the soul (which for all that has so far been said explicitly, is simply a part of the brain), "partly by its own immediate Power, and partly by the help of the Impressions produced by the Senses."[47] Hooke now digresses to calculate whether the brain is large enough to contain all the ideas that are in fact stored in memory. This calculation may have been prompted by a wonder of Boyle's in the previous year, who noted that

> the way whereby the Rational Soul can exercise any power over the humane body, and the way, whereby the Understanding and the Will act upon one another, have not yet been intelligibly explain'd by any. And the like I may say about the *Phaenomena* of the *Memory*. For 'tis a thing much more fit to be admired, than easie to be conceived, how in so narrow a compass, as part of a Human Brain, there should be so many distinct Cells or Impressions as are requisite.[48]

At any rate, Hooke, clearly influenced by his work with microscopes, produced a calculation to show that there was indeed more than enough room.[49] Hooke had earlier noted[50] that in a "*sensible* Point … there may be … many Millions of distinct Particles …." Moreover, he noted, nicely anticipating contemporary allometric studies, the case may well be similar with respect to time:

> I do not at all doubt but that the sensible Moments of Creatures are somewhat proportion'd to their Bulk, and that the less a Creature is, the shorter are its sensible moments, and that a Creature that is a hundred times less than a Man may distinguish a hundred moments in the time that a Man distinguishes one. For when I hear a Fly moving his Wings to and fro so many times, with such a Swiftness as to make a Sound, I cannot but imagine, that that Fly must be sensible of and distinguish at least 3 Moments

---

MSS notes on "The Mechanical Production of Light" are transcribed in Boyle 1999–2000, 14:5–54. See also Golinski 1989.

[45] Hooke 1705, 141.

[46] Hooke 1705, 141.

[47] Hooke 1705, 142–3.

[48] *Advices in Judging of Things said to Transcend Reason*, the 3rd advice, or rule, Boyle 1999–2000, 9:403, published 1681.

[49] Hooke 1705, 143.

[50] Hooke 1705, 134.

> in the time that it makes one of those Strokes with his Wings, for that it is able to regulate and guide it self by the Motion of them. And the like may be said for the quick Motions of other lesser Creatures. So that many of those Creatures that seem to be very short lived in respect of Man, may yet rationally enough be supposed to have lived, and been sensible of and distinguished as many Moments of time as a Man; because within that space of time it has lived, it has had as many distinct Moments of time, and has had as many distinct Differences of Moments, as a Man hath in the Age that he lives. But this only by the by.[51]

7. When the soul produces these ideas ("which I will call little Images") in the "Center of the Repository," it pushes those already present "further into the Repository," and hence the earlier images become further and further away from the centre. This model also allows us to see how, Hooke suggests,[52] we can associate *pastness* with memories, and indeed order them temporally, as well as letting us see how associations spring up among our memories, since those stored in nearby parts of the storage area will influence one another reciprocally.[53] It also accounts for defective memories, and loss of memory, since these images "being material, and so subject to change, … may … be in time alter'd, and sometimes quite lost."[54]
8. But how is it that we are *conscious* of these "little Images"? The soul, centred in the brain, is also the centre of consciousness. Here Hooke offers an analogy between the sun in the external world, and the soul in the neurophysiological realm, an analogy which is, perhaps, not totally compelling. It may be relevant to notice that Hooke was unconvinced by Rømer's measurement of the speed of light[55]:

> this Propagation of Light which is immense, is (in all Probability, and as far as Experiments, Observations and Reasons can assist us) infinitely swift: Or we may say,

[51] Hooke 1705, 134. May and Marten 1983, p. 11, point out that "The fastest physical action of any organism yet recorded is the wing-beat of a common midge (Forcipomyia). It normally beats its wings 57,000 times a minute, but is capable of increasing to a rate of 133,000 times a minute, which represents a muscular cycle of contraction & expansion in 0.00045, or 1/2218th of a second. The fastest muscular cycle in human beings is the blink of an eye, which takes about 1/25 second." (See also Pringle 1949.) Mordenti 1985 notes (with further references) that "When physiologic events in different mammals are measured by biological clocks, they occur in equivalent physiologic time. This … synchronism of time between species is demonstrated by the observation that each mammal lives for approximately the same number of heart beats or breath cycles. Thus, the life span of an elephant and a mouse is the same when measured with a biological clock (i.e., heartbeats), although their life spans vary significantly when measured in years. (887)" See also Brody 1945 and Calder 1981. After Hooke's suggestion, and less plausibly, Bishop Berkeley (1685–1753) suggested that the same holds true for individual humans (see MacIntosh 1978).

[52] Hooke 1705, 144.

[53] For a discussion of Hooke's place with respect to associationism in the history of psychology see Brooks 1981.

[54] Hooke 1705, 144.

[55] Ole Rømer, 1644–1710, published, in 1675, data that showed the speed of light to be finite without, however, calculating a value from his data. (Apparently the first calculation was by Huygens in the *Traité de la lumière*, written in 1678, published 1690.) There is a helpful short chronology of Rømer's work at http://www.rundetaarn.dk/engelsk/observatorium/light.htm.

> that the Propagation thereof through the whole vast or immense *Expansum*, as far as we can yet find, is made in a Point or Instant of time; and at the very Instant that the remotest Star does emit Light, in that very Instant does the Eye upon the Earth receive it, though it be many Millions of Millions of Miles distant, so that in Probability no time is spent between the emitting and the Reception; for with this agrees all the Experiments that have been thought of for this purpose, and no one has yet proved it temporary, though many ways have been thought of for that purpose: And though the ingenious Monsieur *Romer* pretends to have found a way, by which he hath experimentally proved, that this Propagation is not instantaneous but temporary, and so there is somewhat of time spent in the Passage of Light, from the illuminating Object to the Eye or Body enlightned, yet if we examine his Experiment a little more considerately we may find reason to doubt, whether he hath from these grounds sufficient to make such a Conclusion.[56]

So, says Hooke, let's consider the action of the sun. Since it radiates throughout the entire world it may be taken to be, instantly, both affecting and affected by, every point in the universe. Now, "if there were Understanding in the Sun it self" it would thereby gain knowledge of everything its rays impacted.[57] Moreover, if we add to this light rays analogy the fact that the sun also affects things in the world gravitationally, we may consider the soul, centred in the brain as the sun is in the world, able both to affect, and be affected by, the images stored in the brain. "So … the Soul forms to it self a Microcosm, or Picture of the Macrocosm, in which it radiates, and is sensible of every thing contain'd therein, in the same manner as the Sun in the Macrocosm."[58]

But isn't there a drawback to this analogy? For the sun interacts *physically* with the surrounding world, while the soul has the drawback of being incorporeal, so how, exactly, does this interaction occur? On this matter Hooke is charmingly straightforward, "I cannot conceive how the Soul, which is incorporeal, should move and act upon the Ideas which are corporeal, or how those on the other side should … re-act upon and influence the Soul." But, he adds, "I am assured, that such Effects are performed … and without them, neither the Sensation, Cognition,

---

56 Hooke 1705, 77–8. Hooke repeats his dubiety in a number of further remarks. See, e.g., Hooke 1705, 99–100, 108, 130.

57 Hooke 1705, 147. On this analogy see further Henry 1989.

58 Hooke 1705, 147. Both Boyle and Newton had similar views about God's influence in the universe: "In reference to the whole universe, and the creatures it comprises, God may be in some measure resembled by the Magnet, that sustains and pervades, and governs or gives their due dispositions, to the pieces of steel it's Influence reaches to," said Boyle (BP 1990, 4:78; Boyle 2006, §2.2.53, p. 158), and Boyle's younger contemporary Newton took a similar position: "Those ancients who more rightly held unimpaired the mystical philosophy as Thales and the Stoics, taught that a certain infinite spirit pervades all space *into infinity*, and contains and vivifies the entire world. And this spirit was their supreme divinity, according to the Poet cited by the Apostle. In him we live and move and have our being (Dobbs 1991, p. 234)." [Boyle has a similar reference, Boyle 2006 §2.2.47, p. 156&n. The biblical references are to Acts 17.23 and 17.28. Diogenes Laertius, discussing the origin of this and similar altars, identifies the poet as Epimenides (Laertius 1925 1.109–110, pp. 1.114–117)]

Remembring, nor Ratiocination, could be performed; all of which are plainly the results of the conjunct Influences of the Soul, and the Ideas of Bodies placed within the Repository or Sphere of its Activity."[59]

From other seventeenth century writers one might expect, here, an invocation of God's power to sidestep the need for explanation, but Hooke is not given to invoking the deity. As Steven Shapin remarks, "God … is elusive in the published works of Robert Hooke."[60] So we do not have any account of *how* an immaterial soul might interact with Hooke's "mechanical and sensible Figure and Picture" of the mind, nor any suggestion that there is a divine explanation of the mystery, but we do have a straightforward material model of our mental goings-on.

Hooke, then, has offered what he takes to be a plausible analogy of the way in which the brain-centred soul's mechanical activities constitute thinking. And here Hooke stops. "Here," says Richard Waller, the early eighteenth century editor of Hooke's posthumous works, "our Author leaves off, nor can I find, ever reassumed this Subject." Waller then notes, somewhat nervously, "possibly some Persons may imagine that the … Explication … is too mechanical, and tends to the making the Soul a material Being." That is not the case, he assures us. Nonetheless he adds at once: "I hold my self not in the least obliged to defend or maintain any of his Opinions or Discourses."[61]

Hooke, earlier, had similarly disavowed a materialistic doctrine. His views were first offered to the Royal Society on June 21, 1682, when he "read a long discourse, being the substance of three lectures, which he had missed the reading of at two last meetings, concerning the means, how the soul becomes sensible of time, explaining the organ of memory, and its use for retaining and producing ideas therein stored up." A week later a number of members appeared who had not been at the previous meeting, and "it was desired by them, that Mr. *Hooke* should read the same again, which he accordingly did. After which some objecting, that this discourse seemed to tend to prove the soul mechanical, Mr. *Hooke* answered, that no such thing was hinted, or in the least intended in it; it being only designed to show, that the soul forms for its own use certain corporeal ideas, which it stored up in the repository or organ of memory."[62]

Despite Hooke's disclaimer, it is clear that nothing in his account required the mind, or soul, to be immaterial. I conclude that it is at least possible that Hooke was a (closet) materialist.

---

[59] Hooke 1705, 147.

[60] Shapin 1989, p. 277.

[61] Hooke 1705, 148.

[62] Gunther 7:598.

References

Adam C, Tannery P (eds) (1964–1976) Oeuvres de Descartes, 11 vols. Paris
Bacon R (1900) Bridges JH (ed) Opus Maius. London; reprinted Minerva G.m.b.H., Frankfurt/Main, 1964)
Bayle P (1734) Dictionaire Historique et Critique, 5th edn, 5 vols. Amsterdam
Bentley R (1699) The folly and unreasonableness of atheism. Printed by J. H. for H. Mortlock, London. The fourth edition, corrected
Boyle R (1999–2000) In: Hunter M, Davis EB (eds) The works of Robert Boyle,14 vols. Pickering and Chatto, London. Individual works cited by name
Boyle (2006) Boyle on Atheism (transcribed and edited: MacIntosh JJ). University of Toronto Press, Toronto
BP (1990) In: Hunter M (ed) Letters and papers of Robert Boyle. University Publications of America, Bethesda, MD
Brody S (1945) Bioenergetics and growth. Reinhold Publishing, New York
Brooks GP (1981) A note on references by Robert Hooke to the principle of association in an essay on memory. Dalhousie Rev 61:509–512
Brown DJ (2006) Descartes and the passionate mind. Cambridge University Press, Cambridge
Browne T (1646) Pseudodoxia epidemica. Printed by T. H. for Edward Dod, London
Burnet T (1699) Third remarks upon an essay concerning human understanding. Printed for M. Wotton, London, at the Three Daggers in Fleet-street
Burnet T (1989) In: Watson G (ed) Remarks on John Locke with Locke's replies. The Brynmill Press, Doncaster
Calder WA (1981) Scaling of physiological processes in homeothermic animals. Annu Rev Physiol 43:301–322
Clarke E (1968) The doctrine of the hollow nerve in the seventeenth and eighteenth centuries. In: Stevenson LG, Multhauf RP (eds) Medicine, science, and culture. Johns Hopkins, Baltimore, pp 123–141
Clarke DM (2003) Descartes's theory of mind. Oxford University Press, Oxford
Cottingham J (1976) René Descartes, Conversation with Burman (trans: Cottingham J). Clarendon Press, Oxford
Crombie AC (1967) The mechanistic hypothesis and the scientific study of vision. In: Bradbury S, Turner GL'E (eds) Historical aspects of microscopy. Published for the Royal Microscopical Society by W. Heffer & Sons, Ltd., Cambridge, pp 3–112
Cudworth R (1731) A treatise concerning eternal and immutable morality. London. Reprinted Garland, New York, 1976
Dobbs BJT (1991) The Janus faces of genius: the role of alchemy in Newton's thought. Cambridge University Press, Cambridge
Gerhardt CI (1875) In: Gerhardt CI (ed) Die Philosophische Schriften von G. W. Leibniz, 7 vols. Weidmannsche Buchhandlung, Berlin, 1875–1890
Golinski JV (1989) A noble spectacle: phosphorus and the public cultures of science in the early Royal Society. Isis 80:11–39
Grant E (ed) (1974) A source book in medieval science. Harvard University Press, Cambridge, MA
Gunther RT (1930) Early science in Oxford, 15 vols. Oxford University Press, Oxford. Vols. VI and VII: The life and work of Robert Hooke by Richard Waller; Vol. VIII: The Cutler lectures of Robert Hooke
Harvey EN (1957) A history of luminescence from the earliest times until 1900. The American Philosophical Society, Philadelphia
Heath TL (1956) Euclid. *Elements* (translated with Introduction and Commentary by Heath TL, 3 vols.). Dover, New York
Henry J (1989) In: Hunter M, Schaffer S (1989) Robert Hooke: new studies. The Boydell Press, Woodbridge, pp 149–180
Hofmann JE (1974) Leibniz in Paris 1672–1676. Cambridge University Press, Cambridge
Hooke R (1705) Lectures of light; and of comets and gravity. In: Waller R (ed) Posthumous works. London. Reprinted Frank Cass, London, 1971

Jesseph DM (1999) Squaring the circle: the war between Hobbes and Wallis. University of Chicago Press, Chicago
Kant I (1993) Critique of practical reason (trans: Beck LW), 3rd edn. Prentice Hall/The Library of Liberal Arts, Upper Saddle River
Kepler J (1604) Ad Vitellionem paralipomena, quibus astronomiae pars optica traditur. Frankfurt. Reprinted: Culture et Civilisation, Bruxelles, 1968
Laertius D (1925) Lives of eminent philosophers (trans: Hicks RD), 2 vols. William Heinemann/Loeb Classical Library, London
Leibniz GW (1962) In: Robinet A, Schepers H (eds) Nouveaux essais sur l'entendement humain. Academie-Verlag, Berlin
Licetus F (1640) Litheophosphorus Sive De Lapide Bononiensi. Nicolai Schiratti, Bologna
Lindberg DC (1976) Theories of vision from al-Kindi to Kepler. University of Chicago Press, Chicago
Locke J (1975) In: Nidditch PH (ed) An essay concerning human understanding. Clarendon Press, Oxford. References by book, chapter, and section
Loemker LE (1969) Gottfried Wilhelm Leibniz philosophical papers and letters, 2nd edn. D. Reidel, Dordrecht
MacIntosh JJ (1978) Berkeley's views on time. In: Jarrett CE, King-Farlow J, Pelletier FJ (eds) New essays on rationalism and empiricism. Canadian Association for Publishing in Philosophy, Guelph
MacIntosh JJ (1983) Perception and imagination in Descartes, Boyle and Hooke. Can J Philos 13:327–352
MacIntosh JJ (1992) Robert Boyle's epistemology: the interaction between scientific and religious knowledge. Int Stud Philos Sci 6:91–121
Malpighi M (1689) De Structura glandularum conglobatarum. Richardum Chiswell, London
May J, Marten M (1983) Animal oddities. Book Club Associates, London
Möbius AF (1885) Über eine neue Behandlungsweise der analytischen Sphaerik. Leipzig. Reprinted: Möbius. Gesammelte Werke, 4 vols.
Mordenti J (1985) Forecasting cephalosporin and Monobactam antibiotic half-lives in humans from data collected in laboratory animals. Antimicrob Agents Chemother 27.6:887–891
Oldenburg H, Rupert Hall A, Boas Hall M (eds and trans) (1965) The correspondence of Henry Oldenburg, 11 vols., vols 1–9: The University of Wisconsin Press, Madison, 1965–1973; vols 10–11: Mansell, London, 1975, 1977; vols 12–13: Taylor & Francis, London, 1986
Olden-Jørgensen S (2009) Nicholas Steno and René Descartes: a Cartesian perspective on Steno's scientific development. In: Rosenberg GD (ed) The revolution in geology from the Renaissance to the Enlightenment, Memoir 203. The Geological Society of America, Boulder, pp 149–157
Olscamp PJ (1965) René Descartes, Discourse on method, optics, geometry and meteorology (trans: Olscamp PJ). Bobbs-Merrill, Indianapolis
Pecham J (1970) John Pecham and the science of optics: Perspectiva Communis, edited with an introduction, English translation, and critical notes by David C. Lindberg. University of Wisconsin Press, Madison
Porter N (1984) Marginalia Locke-a-na. In: Burnet T, Locke J, Porter N (eds) Remarks upon an essay concerning human understanding: five tracts. Garland, New York
Pringle JWS (1949) The excitation and contraction of the flight muscles of insects. J Physiol 108:226–232
Roda A (1999) The discovery of luminescence: the Bolognian Stone. In: Roda A, Pazzagli M, Kricka LJ, Stanley PE (eds) Bioluminescence & chemiluminescence: perspectives for the 21st century. Wiley and, New York, pp 3–10. An online version may be found at http://www.isbc.unibo.it/Files/10_SE_BoStone.htm
Scott JF (1952) The scientific work of René Descartes (1596–1650). Taylor & Francis, London
Shapin S (1989) Who was Robert Hooke? In: Hunter M, Schaffer S (eds) Robert Hooke: new studies. The Boydell Press, Woodbridge, pp 253–285
Yamada T (2009) Hooke-Steno relations reconsidered: reassessing the roles of Ole Borch and Robert Boyle. In: Rosenberg GD (ed) The revolution in geology from the Renaissance to the Enlightenment, Memoir 203. The Geological Society of America, Boulder, pp 107–126

Chapter 5
Joseph Priestley: An Instructive Eighteenth Century Perspective on the Mind-Body Problem

Alan Beretta

The essence of the soul of man and animals is, and always will be, as mysterious as the essence of matter and bodies
(La Mettrie, Treatise on the Soul, 1745).

5.1 Introduction

It is possible to view episodes in the history of science as potentially edifying with regard to questions that confront modern inquiry. That is to say, the way problems of inquiry were addressed in the past can offer guidance to the way similar problems might be addressed now. In this paper, it is argued that the Cartesian mind-body problem was rendered null and void by Joseph Priestley, an eighteenth century chemist, who brought arguments to bear that are relevant to current concerns about phenomena that are imagined to contrast with 'body' or 'physical' or 'matter'. Specifically, that is to say, Priestley's arguments provide a perspective that can inform discussions regarding the relation between cognition and brain, including the status of qualia, sometimes referred to as the 'hard' problem.

The paper is organized in three parts. First, the main part, Priestley's approach to the mind-body problem is presented. Here, it will be seen that the mind-body problem presupposes a notion of *physical* that, once reasonable, could no longer be maintained in the light of Newton's gravitational force. There could be no principled physical-mental divide because the prerequisite conception of the term *physical*, post Newton, was no longer serviceable. Recognizing this more than 200 years ago is Priestley's great contribution. Second, fast-forwarding through the history of science, it will be contended that no remotely serviceable notion of *physical* has

A. Beretta (✉)
Department of Linguistics, Michigan State University, East Lansing, MI 48824, USA
e-mail: beretta@msu.edu

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_5, © Springer Science+Business Media Dordrecht 2014

ever been resurrected since,[1] so the Newtonian force can, with hindsight, be seen as a first nail in the coffin of physical (or mechanical) explanation. Third, it will be mooted that Priestley's arguments are applicable to current questions about both qualia and cognition-brain relations in general.

It is hoped that, whatever one's convictions with respect to the issues raised, it will be appreciated that Priestley has important things to say where the relation of minds and brains is at issue, and that his writings on the topic deserve more than the neglect they are afforded in some recent histories of the mind-body problem, e.g., Wozniak[2] where he is not featured at all.

5.2 Priestley's Approach to the Mind-Body Problem

5.2.1 Cartesian Dualism

Descartes conceived of *physical* as stuff that had extension in space. Further, and more crucially, he adopted the commonsense assumption that for something physical to move, some other physical entity had to effect that movement; matter pushing matter (known as *contact mechanics*) accounted for motion.

Give him extension and motion, thus understood, and he would give us the universe, as he famously put it. And he did give us the universe, almost. He never did have any idea what the physical explanation of mind might be, but he did come up with a physical explanation of everything else in nature. All phenomena in the universe obeyed the laws of mechanism. That is, they could be grasped the same way we grasp how a machine works. To take an example familiar at the time, the scientist, Huygens, had conjured up a design for a pendulum clock that was considered an advanced piece of mechanical thinking; all of nature could, in principle, be explained the way the workings of Huygens' clock could be explained. The parts of the clock occupied space, and for one part to move, another part had to cause that movement.

Descartes was able to explain the motion of the heavenly bodies without having to savage the notion of contact mechanics. On his account, "the earth is a globe contained in a fluid and mobile heaven"[3] that he likens to a whirlpool. The celestial fluid "turns like a vortex with the sun at its center" and the planets closest to the sun move faster than those further away, and each of them remains "surrounded by the same parts of celestial matter"; this hypothesis "enables us to understand the movement of the planets with great ease."[4] The 'great ease' of this hypothesis is in no small

[1] A point frequently made by Chomsky, e.g., 1995; indeed, much of the present paper is profoundly influenced by what has been called 'Chomsky's physicalism'.

[2] Wozniak 1992.

[3] Cottingham et al. 1985b, vol. 2, p. 253.

[4] Ibid., p. 253.

measure due to its accord with the idea that consorts so well with common sense that contact is required for movement; planets are in contact with a fluid matter that moves them around in their observed orbits. As North puts it, not only Descartes but most cosmologists at the time permitted themselves the "psychological luxury"[5] that mechanical effects were due to the "action of matter pushing matter."[6]

And thus all of nature could be understood. All except mind. Descartes was thus led to posit two entirely distinct and independent substances: *res cogitans* (thinking stuff – mind) in addition to *res extensa* (stuff that occupies space – that is, body). (See, in particular, the sixth meditation).[7]

However, when Newton proposed action-at-a-distance, or gravity, to account for planetary motion, contact mechanics, common sense, great ease, psychological luxury, and the coherence of the idea of mind-body dualism were all swept away.

But before we consider the demise of Cartesian dualism, it is important to be clear that for Descartes the mind-body problem was a perfectly rational problem to arrive at because it was a direct consequence of trying to understand nature in mechanical terms. The guiding theory, as mentioned above, was that everything in nature could be explained in the same way as a machine might be explained, and only one element – mind – presented itself as an anomaly. The divide between 'body' (all of nature) and mind was thus a *theoretical* divide. This is a key point to recognize because it is not always appreciated that, when the theory altered dramatically, as it did with Newton, this divide was immediately obliterated; and it has never since reappeared as a fault-line in nature in any theoretical understanding. And it was Priestley who first appreciated this with respect to the mind-body problem.

The problem of an apparently unbridgeable divide between body and mind was a reasonable one to formulate, so long as 'body/physical/matter' was understood as it was prior to Newton. But not since, as Priestley's work will enable us to see. While there certainly is an extraordinarily challenging problem in seeking to relate mind and brain, the problem is not the one posed by Cartesian dualism. In order to see why Cartesian dualism is mistaken, let us follow the reasoning that led Joseph Priestley from the very same dualist view to its total rejection.

5.2.2 Priestley's Rejection of Dualism

Priestley relates a story that he started out by assuming the "soul [or mind] to be a substance so entirely distinct from matter, as to have no property in common with it".[8] On first giving the mind-body problem some thought, he began to wonder whether "either the material or the immaterial part of the universal system was

[5] North 1995, p. 360.

[6] Ibid., p. 364.

[7] Cottingham et al. 1985a.

[8] Priestley 1777, p. xi.

superfluous".[9] However, failing to make much progress with this line of thought, he "relapsed into the general hypothesis of two entirely distinct and independent principles in man".[10]

These initial thoughts served to focus Priestley's attention on the subject, resulting in a full blown treatise, *Disquisitions relating to matter and spirit*, in which his conjecture that humans were made up of only one sort of substance was developed into a full-scale argument. In the preface to the *Disquisitions*, he presents his conclusions:

> I am rather inclined to think, though the subject is beyond our comprehension at present, that man does not consist of two principles so essentially different from one another as matter and spirit, which are always described as having no common property, by means of which they can effect, or act upon, each other; the one occupying space, and the other not only not occupying the least imaginable portion of space, but incapable of bearing any relation to it; insomuch that, properly speaking, my mind is not more in my body than it is in the moon. I rather think that the whole of man is of *some uniform composition* [italics mine]; and that the property of perception, as well as the other powers that are termed mental, is the result (whether necessary, or not) of such an organical structure as that of the brain…[11]

'Some uniform composition.' A startling thesis at the time, but the idea that the brain does the mind's work or that the mind is a property of brains was not, in fact, entirely new; La Mettrie had come to the same conclusion 30 years earlier via an examination of comparative neurophysiology.[12] Today, by contrast, it is on the cutting edge of the uncontroversial in neuroscience. However, although Priestley's arguments for some uniform composition in some respects would seem innocuous enough to modern neuroscience, in other respects they may be informative in quite surprising ways. Let us deal swiftly with the relatively innocuous matters, before considering the more intriguing respects in which Priestley's contribution might be considered new (there is no trace of it in La Mettrie, for example), insightful, and instructive.

5.2.3 Mind Never Exists Independently of Matter

The part of the argument that is straightforwardly congenial to current opinion is that minds cannot exist independently of matter:

> the powers of sensation or perception, and thought … have never been found but in conjunction with a certain organized system of matter; and therefore that those powers necessarily exist in, and depend upon, such a system. This, at least, must be our conclusion, till it can be shown that these powers are incompatible with other known properties of the same substance; and for this I see no sort of pretence.[13]

---

[9] Ibid., p. xii.

[10] Ibid., p. xii.

[11] Ibid., p. xiii/xiv.

[12] La Mettrie 1747/1996.

[13] Priestley 1777, p. 26.

The 'organized system of matter' is, of course, the brain,

> thought… is a property of the nervous system, or rather of the brain. Because, as far as we can judge, the faculty of thinking, and a certain state of the brain, always accompany and correspond to one another… There is no instance of any man retaining the faculty of thinking, when his brain was destroyed; and whenever that faculty is impeded, or injured, there is sufficient reason to believe that the brain is disordered in proportion; and therefore we are necessarily led to consider the latter as the seat of the former.[14]

What makes the argument historically significant is that it was widely believed that body, far from being a prerequisite for the existence of a mind, is actually "an obstruction to it".[15] Countering those who believe that the body is merely an encumbrance to the faculty of thought, Priestley reasons:

> to suppose [man] capable of thinking better when the body and brain are destroyed, seems to be the most unphilosophical and absurd of all conclusions. If death be an advantage with respect to thinking, disease ought to be a proportional advantage likewise; and universally, the nearer the body approaches to a state of dissolution, the freer and less embarrassed might the faculties of the mind be expected to be found. But this is the very reverse of what really happens.[16]

Finally, just as physiological damage or decay affects the mind, disorders of the mind can impact the body: "the body is liable to be reciprocally affected by the affections of the mind, as is evident in the visible effects of all strong passions, hope or fear, love or anger, joy or sorrow, exultation or despair".[17]

Priestley's appeal to general observations that as brains go, so do minds, is offered in support of his contention that minds are properties of brains. That the appeal was made in these terms is historically interesting in its own right, since it was not a widely held view, but the issue is regarded more or less as a truism today, so its usefulness to current debates is limited. Let us now turn to an aspect of Priestley's argumentation that might give us rather more to think about.

5.2.4 Matter Never Exists Independently of Forces and Forces Move Matter Without Contact

Critical to the aspect of Priestley's reasoning that we now consider is the idea of forces of attraction and repulsion. Most important of all was Newton's 'action at a distance', the gravitational force, but he also alludes to the work of scientists of his own era, like Boscovich (sometimes considered the founder of atomic physics). He does not mention the increase in understanding of the force that operates on electrical charges, but this is plainly assumed, since he makes frequent reference

[14] Ibid., p. 27.

[15] Ibid., p. 29.

[16] Ibid., p. 29.

[17] Ibid., p. 28.

to the powers of attraction and repulsion; indeed, Priestley himself,[18] along with his contemporaries Cavendish and Robison, had contributed to that growing understanding, even before Coulomb's Law was stated (which was shortly after *Disquisitions* was published).

Priestley advances a claim that the commonsense notion of matter, of the sort that the Cartesian mind-body dualism presupposes, has been completely undermined by insights provided by developments in inquiry into nature. Where previously it had been taken for granted that bodies have properties of extension and solidity, science had now shown that matter, thus understood, did not exist. As he elaborates the claim, solidity was inherent only in *points*. In Boscovich's theory, that Priestley draws from, *points* were like modern protons and neutrons, and these *points* form only a tiny part of atoms. He even conjectures that "for anything we know to the contrary, all the solid matter in the solar system might be contained within a nutshell, there is so great a proportion of void space within the substance of the most solid bodies".[19] More tellingly, matter, as now understood by science, possessed *powers*, forces of attraction and repulsion, which themselves lacked solidity, did not occupy space, and were not inert, but nevertheless, were now the essential properties of the newly conceived matter. This is a very different conception of matter than obtained in the Cartesian worldview. So, what looked and felt like a solid table that had extension in space was really an assembly of 'points' that formed but a tiny fraction of the whole, a whole that was held together in its apparent shape by 'powers' that themselves lacked extension. It is hard to imagine what construal of the term *physical* would include both forces and objects that occupy space.

In Priestley's words, "matter is not the inert substance that it has been supposed to be"; rather, "powers of attraction and repulsion are necessary to its very being".[20]

The power of repulsion acts "at a real, and in general an assignable, distance from what we call the body itself".[21] Returning to the earlier discussion of the hand pressed against a table, supposedly a case of two solids making contact, Priestley commented that "notwithstanding their seeming contact, they are actually kept at a real distance from each other, by powers of repulsion common to them both".[22] Hands interact with tables *without direct contact*.

If we try to imagine entities that do not possess the power of attraction, "a power which has always been considered something quite distinct from matter itself, [then] there cannot be any such thing as matter"[23]; rather, "this foreign property, as it has been called, is in reality absolutely essential to its very nature and being. For when we suppose bodies to be "divested of it, they come to be nothing at all"[24] If we take

---

[18] Priestley 1767.

[19] Priestley 1777, p. 17.

[20] Ibid., p. xxxviii.

[21] Ibid., pp. 4–5.

[22] Ibid., p. 12.

[23] Ibid., p. 5.

[24] Ibid., p. 5.

way attraction, "solidity itself vanishes"; without doubt, if it were not for the attractive force, particles would just fall apart and "be dispersed"[25]

This is now ordinary chemistry. We understand chemical bonds as forces that hold atoms together. Thus, covalent bonds are thought of in terms of attraction and repulsion. To flesh out a commonly used example, we recall that the positively charged nuclei of two hydrogen atoms repel each other, as do the two negatively charged electrons, but each of the nuclei are simultaneously attracted to the same electrons. If the two atoms are the right distance apart from each other, then the attractive forces are stronger than the repulsive forces and the shared electrons form a bond between the two atoms. If there were no forces, then there would be no bonds, no solid structures, nothing beyond the smallest particles that science has discovered that even occupy space.

It is fascinating to learn that Priestley, aware of Boscovich's conjectures at the atomic level, was able to imagine that even at that level, things fall apart and the center cannot hold if there are no forces to preserve their integrity. That is, his argument that there could be no gross bodies if there were no forces applies to the atom as well, since he believes that it, too, must be decomposable into smaller elements themselves held together by forces. As he puts it,

> [the] atom must be divisible, and therefore have parts … and therefore [the parts] must have powers of mutual attraction infinitely strong, or it could not hold together, that is, it could not exist as a solid atom. Take away the power therefore, and the solidity of the atom entirely disappears. In short, it is then no longer matter; being destitute of the fundamental properties of such a substance.[26]

In principle, according to Priestley, there is no difference between tables, or brains, and atoms insofar as they all depend on forces for their very existence:

> For as the large bodies would be dissolved without some principle of union, or some power, internal or external, so the parts of which they are composed would, in similar circumstances, be resolved into smaller parts, and consequently (the smallest parts being resolved in the same manner) the whole substance must absolutely disappear, nothing at all being left for the imagination to fix upon.[27]

5.2.5 Consequences for Mind-Body Problem

What is the status of the mind-body problem in this brave new post-Newtonian world that Priestley confronted? The principal consequence of the emergence of forces was that the term *body* ceased to mean what it had meant before. *Body* now had to take forces into account, and thus was robbed of the meaning that served as a stark contrast with mind. Extension turned out to be a consequence of forces that operate both within and between atoms; and these forces move objects around

[25] Ibid., p. 6.

[26] Ibid., p. 6.

[27] Ibid., pp. 6–7.

without direct contact. Thus the kind of body that was to be contrasted with *res cogitans* no longer obtained. Given this development, Priestley concluded:

> If I be asked how, upon this hypothesis matter differs from spirit… I answer, that it no way concerns me… to maintain that there is any such difference between them as has hitherto been supposed. On the contrary, I consider the notion of the union and mutual influences of substances so essentially different from one another, as material and immaterial substances have been represented, as an opinion attended with difficulties infinitely embarrassing, and indeed actually insuperable.[28]

Priestley considered that it made sense to conceive of a kind of matter that comprised both mind and body, "especially as we know nothing more of the nature of substance than that it is something which supports properties" and these properties "can be anything we please, provided they be not inconsistent with each other". Thus, mind, or "powers of sensation or perception, and thought", could belong to the same substance as one that supports the properties of attraction and repulsion.[29] That is, 'some uniform composition,' a brain-mind.

The motivation for positing mind as some distinct substance from body has vanished: "since the only reason why the principle of thought, or sensation, has been imagined to be incompatible with matter" was based on a concept of matter that has been shown to be mistaken, "the whole argument for an immaterial thinking principle in man… falls to the ground".[30] Given the demise of the obvious contrast of a notion of body with mind, continuing to assume a distinct thinking substance has nothing to recommend it: "If one kind of substance be capable of supporting all the known properties of man; that is, if those properties have nothing in them that is absolutely incompatible with one another, we shall be obliged to conclude… that no other substance enters into his composition".[31]

Priestley began as a commonsense dualist, but by thinking through the consequences of Newtonian forces and further scientific developments in the intervening century, he reasoned his way to a rejection of dualism and towards a concept of matter that included the mental. He had no idea what properties the mental might possess, "we have a very imperfect idea of what the power of perception is… but this very ignorance ought to make us cautious in asserting with what other properties it may, or may not exist".[32] However, he did know that whatever *body* was, possessing forces but not occupying space (other than at some particle level), it was definitely not the kind of stuff that the Cartesian mind-body problem assumed. So, the unification of the mental and the physical would not be across the same divide. No longer was it a divide between body and mind because *body* had changed beyond recognition.

---

[28] Ibid., p. 16.

[29] Ibid., pp. 17, 22.

[30] Ibid., p. 18.

[31] Ibid., pp. 24–25.

[32] Ibid., p. 26.

In Chomsky's evocative phrase, "Newton eliminated the problem of 'the ghost in the machine' by exorcising the machine; the ghost was unaffected".[33]

5.3 The Physical-Nonphysical Divide Has No Status

To posit something in contrast to *physical* presupposes that we know what *physical* means. In Descartes' time, we thought we knew, but with Newtonian action-at-a-distance, it became clear that we did not.

In a sense, of course, we *could* claim that we do know what *physical* is. We *could* say that *physical* is the Priestleyan 'anything we please' that "finds a place in intelligible explanatory theory".[34] That is, it is anything that we have some theoretical understanding of (so long as it agrees with evidence). But this sense of *physical* is not of much use to us either. Certainly, it is not something we can use to create a divide between brain and mind, brain being physical and mind not. After all, we have a theoretical understanding of brain and, many cognitive scientists would affirm, a theoretical understanding of some aspects of mind, both of which agree with evidence. So, to assume that there is a physical-nonphysical divide is to assume, counterfactually, that we have some useful notion of *physical*. (Of course, the terms *physical* and *matter* are still used, for instance when we talk about such things as the 'electron theory of matter', or even the unfathomable 'dark matter', but this is just a vague way of referring to whatever concepts occur in our theories.)

The concept of matter has assumed wildly different forms again and again in the centuries since Newton. It has been forced to accommodate, for example, electromagnetic forces, strong and weak nuclear forces, quantum theory, and potentially, the exotic possibilities held out by string theory. The only divide that makes any sense now with respect to mind and brain is between whatever our best theories of the mental and the neural are.

We have little more idea now than Priestley had of what kind of matter it could possibly be that could accommodate both a theory of an aspect of mind and a theory of brain. We have a theory unification problem, a sort of problem that, after several centuries of experience, is familiar in the advanced sciences.[35] It does not mean that we will solve the problem, but at least now we know what the problem is and what it is not.

As Priestley was able to appreciate, a physical-nonphysical divide has no justification. The term *physical* has no useful status at all; indeed, it has *no* status. All we have are phenomena in the world that present themselves to us, and we try to understand them as well as we can, that is, construct theories about them and test them.

[33] Chomsky 2000, p. 84.

[34] Chomsky 1995, p. 5.

[35] Ibid.

Historically, it has never been easy to accept that physical, or by extension, mechanical explanation is dead. So, let us turn to some examples in the history of science, starting with reaction to Newton.

Newton's contemporaries famously reacted to the idea of a gravitational force with derision. Huygens found it absurd; how could it give rise to the motion of planets if it acted through a vacuum, through etherless, particle-free space? Obviously, there could be no physical mechanism in a vacuum, so the idea was a non-starter. Leibniz dismissed action-at-a-distance as occult; it was just a mathematical story but it had no physical basis and thus lacked even a pretense of a mechanism that would explain motion. Newton himself considered it equally absurd; in his correspondence, he had this to say:

> That gravity should be innate, inherent, essential to matter, so that one body may act upon another at a distance through a vacuum, without the mediation of anything else, by and through which the action and force may be conveyed from one to another, is to me so great an absurdity that I believe no man who has in philosophical matters a / competent faculty of thinking can ever fall into it. Gravity must be caused by an agent acting constantly according to certain laws, but whether this agent can be material or immaterial I have left to the consideration of my readers (cited in Kline).[36]

Nobody ever did find the physical or mechanical basis of gravity. It was proposed as a mathematical explanation and that is what it remains. Galileo had earlier advocated purely mathematical explanations, abandoning mechanical causes, for instance with respect to falling bodies, and Newton had apparently adopted this approach, though with the reluctance and reservations expressed in the above letter. It was not just that mathematics provided a precise means of articulating something that had an understandable physical cause, it was that mathematics was all there was in terms of explanation. It's hard to imagine the shock experienced in the late seventeenth century when the freshness was still upon the idea of a force, especially as there had previously been such optimism that nature would turn out to be explicable in just the way it presented itself to our common sense. Bacon anticipated that all of nature's secrets would be revealed to us; it was just a matter of time. But Newton put paid to that, once and for all. As the late mathematician, Morris Kline, so concisely and clinchingly put it, "Newton's *Principles* is an epitaph to physical explanation".[37] The stunning finality of Kline's remark matches perfectly the finality of all hopes for a commonsense physical understanding of the world; Newton's force had extinguished them.

Leibniz, Huygens, and others confronted by a new world of theoretical advances without physical mechanism were dismayed. But as we fastforward to the nineteenth century, we will see that some leading scientists continued to be dismayed. That is, with new theoretical advances in electromagnetism, also without physical mechanism, scientists reacted just as Huygens and Leibniz did. Which is a measure of just how hard it is to accept the demise of physical explanation.

---

[36] Kline 1980, pp. 55–56.

[37] Ibid., p. 57.

With Maxwell's equations and the development of electromagnetic theory, the desire for physical explanation was utterly confounded. Electric fields, magnetic fields, electromagnetic waves, all resisted physical understanding and continue to do so to this day. As Kline describes it, "Electromagnetic theory is entirely a mathematical theory illustrated by a few crude physical pictures… Radio waves and light waves operate in a physical darkness… the best we can do… is to fit inadequate physical pictures to mathematical facts;" Faraday wrote to Maxwell asking if he could express his equations in "common language as fully, clearly, and definitely as in mathematical formulae?… translating them out of their hieroglyphics." But Maxwell could not; nor could anyone. Lord Kelvin (William Thomson) was distinctly unimpressed: "I never satisfy myself until I can make a mechanical model of a thing. If I can make a mechanical model I can understand it. As long as I cannot make a mechanical model all the way through I cannot understand; and that is why I cannot get the electromagnetic theory." As Helmholz noted, "in Maxwell's theory an electric charge is but the recipient of a symbol;" electrons, fields, and waves, are just labels for entities that appear in formulas.[38]

But even the incorporeal nature of electromagnetism could not prepare us for quantum theory. In Heisenberg's words, "I repeated to myself again and again. Can Nature really be as absurd as it seemed to us in these atomic experiments?" (cited in Kline).[39] Weirder and more bewildering than anything that had gone before, it taunts us with the impossibility of a physical explanation. It has become an iconic example of just how removed from commonsense understanding science has become: On a discussion of the behavior of electrons on a recent PBS radio program, a physicist was interviewed thus: "So, could my producer be both in her office and out of her office at the same moment?" The physicist responded: "Yes, if she were very small and loosely coupled to her environment!" We know we are not going to grasp even the rudiments of this bizarre theory if we demand it be reduced to a physical or mechanical model; hence Feynman's sustained preparation of his readers, entreating them to abandon common sense, so that they might be ready to "accept some very bizarre behavior" in quantum electrodynamics.[40]

The advanced sciences are "ghost fields", in Kline's apt phrase; he continues, "Modern science has been praised for eliminating humors, devils, angels, demons, mystical forces, and animism by providing rational explanations of natural phenomena. We must now add that modern science is gradually removing the intuitive and physical content, both of which appeal to the senses; *it is eliminating matter*; it is utilizing purely synthetic and ideal concepts such as fields and electrons about which all we know are mathematical laws" (italics mine).[41]

One can readily understand why Huygens and his contemporaries might demur at a proposal that lacked a mechanical basis. They were the first to be confronted with the crushing realization that instead of directly apprehensible mechanical

[38] Kline 1985, pp. 143–146.

[39] Ibid., p. 181.

[40] Feynman 1995, p. 119.

[41] Ibid., p. 119.

explanation, they would thereafter have to accept absurd, occult, bizarre, theory. They would have to accept that the natural sciences are ghostly and ethereal. Two centuries later, Lord Kelvin and others appeared to go through the same torment; indeed, Kelvin could not yet accept a non-mechanical explanation even for gravity. In the twentieth century, Heisenberg's reaction was stunned acceptance that the atomic world really did seem to be as absurd as he and his contemporaries were discovering it was; by this time, physicists were, as Feynman put it, used to it: "They've learned to realize that whether they like or don't like a theory is *not* the essential question. Rather, it is whether or not the theory gives predictions that agree with experiment" no matter "how absurd from the point of view of common sense".[42] As for mechanical models, "the more you see how strangely Nature behaves, the harder it is to make a model that explains how even the simplest phenomena actually work. So theoretical physics has given up on that"[43]; "there are no 'wheels and gears' in our theories of nature".[44]

Ghost fields, theories without physical mechanism, without 'wheels and gears', are the currency that modern inquiry has to deal with, and this is true of neurobiology, too. Granted, if we open any neurobiology textbook, we will see reference to mechanism after mechanism, apparent wheels and gears on every page. But this is just a useful way of characterizing structures and processes once all the ghosts are assumed. Consistent with Priestleyan reasoning, absent forces, there would be no neuronal structures, no potentials of any sort, no binding of neurotransmitters and receptors; in short, no brains.

5.4 Priestleyan Reasoning and Current Discussions of Qualia

How to explain qualia, our sense of conscious experience, is a vexed problem in cognitive science. One of the more well known attempts to grapple with the problem has been made by Chalmers.[45] On Chalmers' account, there is a chasm between the physical and experience. On the physical side of the divide, explanation of structures and functions is possible. To explain the functions, all we need do is specify the mechanism. "Purely physical explanation is well-suited to the explanation of physical *structures*, explaining macroscopic structures in terms of detailed microstructural constituents; and it provides a satisfying explanation of the performance of *functions*, accounting for these functions in terms of the physical mechanisms that perform them".[46] So, to explain functions like language, or memory, or

[42] Ibid., p. 10.

[43] Ibid., p. 82.

[44] Ibid., p. 78.

[45] Chalmers 1995.

[46] Ibid., p. 208.

learning, the task is to specify the neural mechanisms that are responsible for these functions. On the other side, the nonphysical side, we have experience. "Experience" as Chalmers points out, "may *arise* from the physical, but it is not *entailed* by the physical",[47] so a purely physical account will inevitably fail to account for it. That is to say, even if all the physically-dependent functions and mechanisms were understood, experience would remain unexplained. The solution to the problem that Chalmers offers is to treat experience as a fundamental entity in the universe, underivable from anything else, on a par with Maxwell's electromagnetic force. This way, it is hoped, it may prove possible to devise some sort of theory of experience.

Priestley arrived at a view that the physical-nonphysical divide was untenable (with respect to the mind-body problem) and, as argued above, no physical-nonphysical divide has emerged in natural inquiry in the centuries since (attendant upon the loss of any serviceable idea of what counts as physical). In the light of Priestley's perception and its confirmation in the intervening period, the obvious question to ask about Chalmers' discussion of experience is what is meant by *physical*.

Given the references in the discussion to 'physical theory', it seems clear that what is intended is something along the lines of 'whatever science has come up with so far', which is to say, our best current theories supported by experiment. Experience may 'arise' from these theories but not be 'entailed' by them. Recall that minds, to Priestley, were mysterious ("we have a very imperfect idea of what the power of perception is"); every bit as mysterious as conscious experience is now; which is to merely to say, we had no theory of any aspect of mind in the eighteenth century that had any evidential basis, just as we have no theory of qualia now that has any evidential basis. Recall also that Priestley added, "but this very ignorance ought to make us cautious in asserting with what other properties it may, or may not exist". If we have no more than a vague, pretheoretical idea of what conscious experience might be, it is far from obvious that the claim that it cannot be part of the physical world (understood simply as the best evidence-based understandings humans have been able to come up with to date) is warranted. It might be that it *is* in fact entailed by current theories in the advanced sciences, but we cannot tell whether it is or it is not so long as we have no theory of experience to compare them with. All we have today is the way experience presents itself to us, which might bear little relation to reality. After all, we have a long history of having to accept theories that tell us that the world as it presents itself to us is not the way the world is.

But even if we assume that a theory of conscious experience, if we can ever devise one that has experimental support, turns out to be something that does not follow from (is not entailed by) current scientific understandings, all that we would have then is something else that is familiar in inquiry: a problem of how to integrate theories that do not seem to have a property in common. Prior to Newton, who could see that it would be possible to integrate terrestrial and celestial motion? Prior to Maxwell and his contemporaries, who imagined that electricity, magnetism, and light would be integrated?

---

[47] Ibid., p. 208.

Theories in the advanced sciences might change dramatically such that experience falls out as a consequence. Specifically, some development in neurobiology, perhaps one as dramatic as the quantum revolution, might permit us to see experience as a consequence. Chalmers dismisses any appeal to a future neurobiology "It is difficult to imagine what a proponent of new neurophysiology expects to happen, over and above the explanation of further cognitive functions. It is not as if we will suddenly discover a phenomenal glow inside a neuron!"[48] Yet one wonders how is it possible to assert, lacking a theory, that experience is not a cognitive function? Or, that some future neuroscience will not be able to explain more than functions? Furthermore, discovering 'a phenomenal glow inside a neuron' is no more ridiculous than proposing a force that operates across a vacuum. To Newton's contemporaries, it was as if "Newton had stated that the sun generates in the planets a quality which makes them describe ellipses" (Dijksterhuis,[49] cited in Chomsky),[50] as if he had discovered a phenomenal glow inside planets. It was precisely because this 'glow inside planets' wiped out all previous conceptions of *physical* that Priestley realized that whatever mind might be, it was not clear that it contrasted with whatever body had become.

Regarding the proposal to consider experience as a fundamental entity on a par with electromagnetism, it should perhaps be noted that Maxwell's massive achievements were possible because of a plethora of experimental results and mathematical insights; he had 'cause and will and strength and means' to come to a theory of electromagnetic forces. The motivation, in the case of experience, does not seem to be on a par, to put it mildly. Priestley argued against the idea of positing a distinct thinking substance because "true philosophy [drawn from Newton's rules of philosophizing] will not authorize us to multiply causes, or kinds of substance, without necessity, [and so it] will forbid us to admit of any such substance".[51] Lacking any necessity to posit a distinct experiencing substance (since we lack a theory, since future theories of experience and brains may be compatible, and since phenomenal glows inside neurons are not more exotic or bizarre than what is contained in other theories), about all that is accomplished by the proposal is to multiply causes.

5.5 Final Remarks

It could not have been easy, writing in the eighteenth century, to draw the inferences from the Newtonian revolution for mind-body relations that Priestley did. So far as I know, in this respect, he did the original thinking. Recognizing that there was no principled divide between entities characterized as physical and those considered nonphysical was an important insight, when he first articulated it, for any effort to

[48] Ibid., p. 207.

[49] Dijksterhuis 1986.

[50] Chomsky 1995.

[51] Priestley 1777, p. 24.

frame the problem in a manner that was responsive to developments in scientific understanding. In this chapter, I have tried to make the case that it is also important now as attempts to come to terms with the far reaches of what we think of as cognitive unfold. At the very least, it might be regarded as rather surprising that his arguments have not received a more central place in discussions of brain-cognition relations.

References

Chalmers DJ (1995) Facing up to the problem of consciousness. J Conscious Stud 2:200–219
Chomsky N (1995) Language and nature. Mind 104:1–61
Chomsky N (2000) New horizons in the study of language and mind. Cambridge University Pres, Cambridge
Cottingham J, Stoothoff R, Murdoch D (eds) (1985a) The philosophical writings of Descartes: vol 1. Cambridge University Press, Cambridge
Cottingham J, Stoothoff R, Murdoch D (eds) (1985b) The philosophical writings of Descartes: vol 2. Cambridge University Press, Cambridge
de La Mettrie JO (1747/1996) Machine man and other writings. Thomson A (trans and ed). Cambridge University Press, Cambridge
Dijksterhuis EJ (1986) Mechanization of the world picture. Princeton University Press, Princeton
Feynman R (1995) QED: the strange theory of light and matter. Princeton University Press, Princeton
Kline M (1980) Mathematics: the loss of certainty. Oxford University Press, New York
Kline M (1985) Mathematics and the search for knowledge. Oxford University Press, New York
North J (1995) The Norton history of astronomy and cosmology. Norton & Co., New York
Priestley J (1767) The history and present state of electricity, with original experiments. Printed for Dodsley J, Johnson J, & Cadell T, London
Priestley J (1777) Disquisitions relating to matter and spirit. Kessinger, Whitefish
Wozniak RH (1992) Mind and body: René Descartes to William James. National Library of Medicine/American Psychological Association, Bethesda/Washington, DC

Chapter 6
Reflections of Western Thinking on Nineteenth Century Ottoman Thought: A Critique of the 'Hard-Problem' by Spyridon Mavrogenis, a Nineteenth Century Physiologist

George Anogianakis

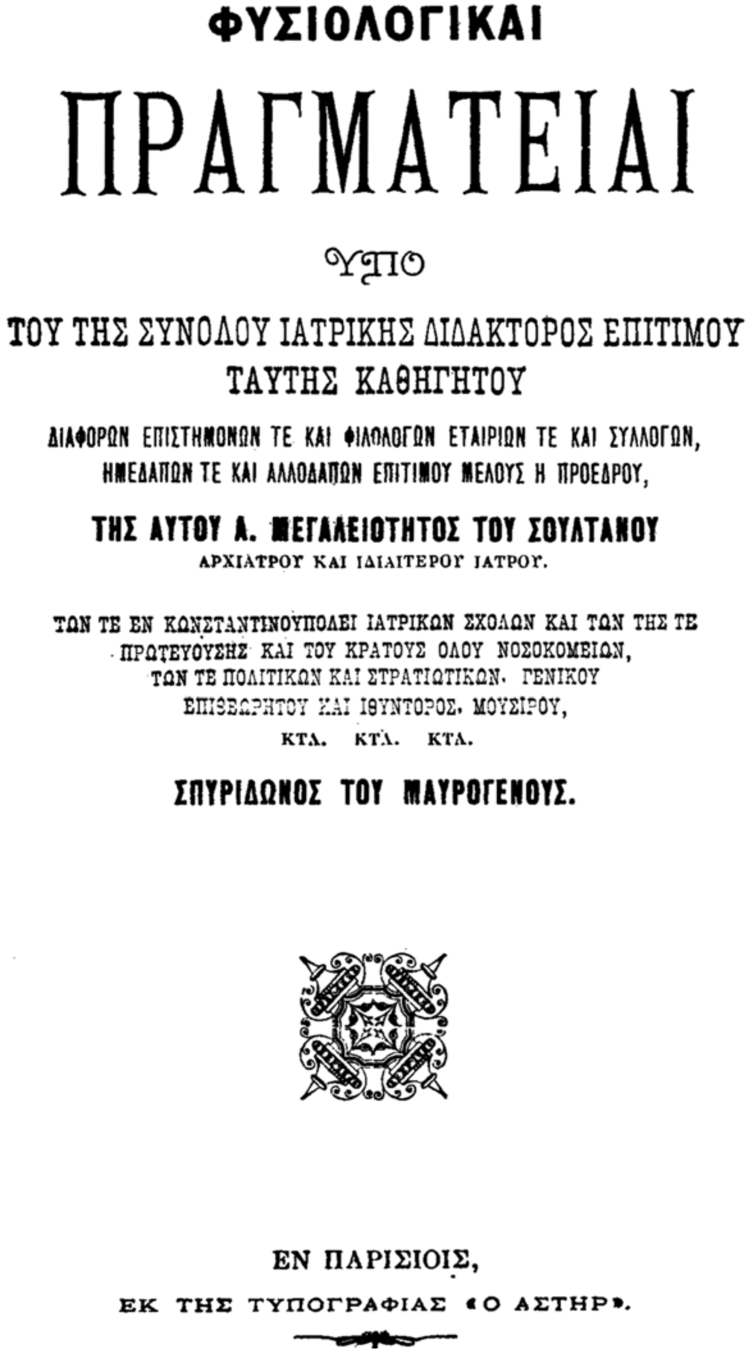
6.1 Introduction

In his attempt to educate the Ottoman Empire's public about the achievements of Western Thought in the nineteenth century, Spyridon Mavrogenis wrote an 85 page treatise about "Life and Soul" which is the subject of this review. To understand Mavrogenis' philosophy one must have in mind that, at the time he wrote this treatise, he was one of the most influential professors of Medicine in the Istanbul medical school where he taught the course in internal medicine. In his treatise he clearly presents the then state-of-the-art in Physiology, makes a clear distinction between Physiology and Psychology as to their methodologies and correctly describes the then current State of the "Brain – Mind problem". At the same time he takes a calculated position as not to disturb the powers of his time, the Sublime Porte (i.e., the Sultan) and the Eastern Orthodox Church! He clearly denounces materialism when it comes to the position materialists take on religion, specifically decries those who think of thoughts as "stuff produced by glands" and he admonishes his readers not to trust that the human mind (irrespective of whether idealistic or materialistic in its essence) can go deep enough into the mysteries of the physical universe to answer questions about its ultimate essence.

It is the impression of the author that, ever since the Middle Ages, the West has dominated modern science and, rather chauvinistically, it considers Eastern thinking as an intellectual activity of inferior caliber. However, problems like the hard-problem do not exclusively belong to Western thought. They are universal problems and they must be dealt with as such. On the other hand, there are very few treatises of Middle Eastern origins that deal with the nineteenth century views on the hard-problem and are readily available to the West. "Life and Soul," the 85

G. Anogianakis (✉)
Department of Experimental Physiology, Faculty of Medicine,
Aristotle University of Thessaloniki, 54124 Greece
e-mail: Giorgos.anogeianakis@gmail.com

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_6, © Springer Science+Business Media Dordrecht 2014



ΦΥΣΙΟΛΟΓΙΚΑΙ

ΠΡΑΓΜΑΤΕΙΑΙ

ΥΠΟ

ΤΟΥ ΤΗΣ ΣΥΝΟΔΟΥ ΙΑΤΡΙΚΗΣ ΔΙΔΑΚΤΟΡΟΣ ΕΠΙΤΙΜΟΥ
ΤΑΥΤΗΣ ΚΑΘΗΓΗΤΟΥ

ΔΙΑΦΟΡΩΝ ΕΠΙΣΤΗΜΟΝΩΝ ΤΕ ΚΑΙ ΦΙΛΟΛΟΓΩΝ ΕΤΑΙΡΙΩΝ ΤΕ ΚΑΙ ΣΥΛΛΟΓΩΝ,
ΗΜΕΔΑΠΩΝ ΤΕ ΚΑΙ ΑΛΛΟΔΑΠΩΝ ΕΠΙΤΙΜΟΥ ΜΕΛΟΥΣ Η ΠΡΟΕΔΡΟΥ,

ΤΗΣ ΑΥΤΟΥ Α. ΜΕΓΑΛΕΙΟΤΗΤΟΣ ΤΟΥ ΣΟΥΛΤΑΝΟΥ
ΑΡΧΙΑΤΡΟΥ ΚΑΙ ΙΔΙΑΙΤΕΡΟΥ ΙΑΤΡΟΥ.

ΤΩΝ ΤΕ ΕΝ ΚΩΝΣΤΑΝΤΙΝΟΥΠΟΛΕΙ ΙΑΤΡΙΚΩΝ ΣΧΟΛΩΝ ΚΑΙ ΤΩΝ ΤΗΣ ΤΕ
ΠΡΩΤΕΥΟΥΣΗΣ ΚΑΙ ΤΟΥ ΚΡΑΤΟΥΣ ΟΛΟΥ ΝΟΣΟΚΟΜΕΙΩΝ,
ΤΩΝ ΤΕ ΠΟΛΙΤΙΚΩΝ ΚΑΙ ΣΤΡΑΤΙΩΤΙΚΩΝ. ΓΕΝΙΚΟΥ
ΕΠΙΘΕΩΡΗΤΟΥ ΚΑΙ ΙΘΥΝΤΟΡΟΣ. ΜΟΥΣΙΡΟΥ,
ΚΤΛ. ΚΤΛ. ΚΤΛ.

ΣΠΥΡΙΔΩΝΟΣ ΤΟΥ ΜΑΥΡΟΓΕΝΟΥΣ.

ΕΝ ΠΑΡΙΣΙΟΙΣ,
ΕΚ ΤΗΣ ΤΥΠΟΓΡΑΦΙΑΣ «Ο ΑΣΤΗΡ».

Fig. 6.1 Front page of Treatises in Physiology, by Spyridon Mavrogenis. The book was published by a Greek publisher ("Astir") in Paris France. Although the date of publication is not mentioned, the book was apparently published in the last few years of the nineteenth century page treatise of Mavrogenis, which is part of a volume titled "Treatises in Physiology"[1] (Fig. 6.1), is one of these few. Written in a very archaic Greek idiom (Katharevousa),[2] almost as distant from Modern Greek as the Septuagint translation of the Hebrew bible is, it was directed to the Greek audience of late nineteenth and early twentieth century Istanbul. Still, coming from an author whose qualifications include being the personal physician of the Sultan, we cannot escape its authority when it comes to presenting the locally prevailing ideas and, sometimes, attitudes. Therefore, in order to understand the locally prevailing intellectual climate and to be able to correctly interpret Mavrogenis' views, a short description of Islamic and Turkish medical science is first presented, before presenting a review of the Mavrogenis text itself. In addition, given that in building their imperial administrative system, the Ottomans assimilated many Byzantine administrative practices, a brief genealogy of the Mavrogenis family is included so that the process, by which an "infidel" could ascend to such an important position in an Islamic State, is understood. Finally it is noted that all ideas, statements or descriptions of the work of philosophers and scientists mentioned in the text, are Mavrogenis' own interpretation and no attempt has been made by the author to present these person's work in their own words.

6.2 Ottoman Contributions to Medical Science

"The medieval Islamic world, from Central Asia to the shores of the Atlantic, was a world where scholars and men of learning flourished"

– From a speech by HRH The Prince of Wales titled "Islam and the West," at the Oxford Centre for Islamic Studies, The Sheldonian Theatre, Oxford, 27th October 1993

The history of Islam starts with the Hijra in 622 CE, when Prophet Muhammed moved from Mecca to Medina. By 650 CE, Islam had conquered Syria, Egypt, Mesopotamia, and Persia. The local civilizations that they met, superior to their own, were assimilated and they contributed substantially to the development of Muslim civilization. By 750 CE the rise of the Abbasid Caliphate, which paid great attention to knowledge and culture, ushered a cultural and scientific era onto the Islamic world. Both Indian and Greek science and medicine were discovered by Islam while Greek, Sanskrit, Syrian and Pahlavi literature was systematically translated into Arabic. For our purposes, however, the main medieval Islamic cultural achievement was the organization and systematization of the educational process and the dissemination and popularization of learning. Learning was

[1] Spyridon Mavrogenis. Treatises in Physiology, ca. 1895–1900, "Astir" Publishers, Paris France.

[2] Katharevousa (Καθαρεύουσα) literally means "puristic language". It is a conservative form of the Modern Greek language conceived in the early nineteenth century as a compromise between Ancient Greek and Demotic of the time. Through the last quarter of the twentieth century, it was widely used mainly for official purposes, though seldom in daily language.

considered essential for the development of every Muslim, as evidenced by the number of public libraries and schools that existed in Islamic lands. The promotion of education was one of the duties of the Islamic State. Although many medieval Islamic thinkers of the time interpreted "knowledge" as to exclude mathematical, physical and natural sciences, these, so called "secular sciences" proved to be an important part of the Islamic educational system.

The introduction of paper manufacturing had as great an impact on education and dissemination of knowledge across the Islamic world as Guttenberg's invention of printing had on European education. Paper may have also facilitated contact between Europe and the Islamic World, while the Islamic popularization of education may have contributed substantially to this end. The many parallels, between the Islamic madrasa and the late medieval European university, raise the possibility for such a contact and intellectual exchange to have indeed taken place.

The madrasa system came into being during the period of Turkish Seljuk rule. The Seljuk dynasties were the founders of the earliest madrasa schools as well as of many of the libraries that sprang up across the medieval Islamic World during their rule. The patronage and encouragement of scientists and their work by sultans and viziers (prime ministers) was of cardinal importance for the production and continuation of scientific work. The utility of a branch of knowledge was considered as the most legitimate criterion for dispensing such patronage. Such a criterion, sometimes, led to substantial support for disciplines like astrology and alchemy (which were perceived as divination methods rather than as pseudoscience). However, medicine enjoyed, both popular and royal, lavish support. In fact, it can be said that not only medicine but every basic secular science enjoyed increased support during Turkish Seljuq rule.

Apart from lavish support, the development of Islamic Physiology and Medicine benefited immensely from the emigration of scientists and physicians from the predominately fundamentalist Christian States of the Middle East and their settlement to the more religiously tolerant Arab and, later, Turkish lands. For example, even during its pre-Islamic era, Jundisapur[3] is said to have owed its fame and importance (as center of medical excellence) to Nestorian physicians who were forced to leave the Eastern Roman Empire because of their religious views. Although, recently, serious doubts have been raised as to the existence of the Jundisapur hospital, it would be natural for Nestorian physicians to seek the protection of close-by Iran and to bring along their trade. Tradition has it that the Jundisapur hospital (or some equivalent institution in the area, if we accept the conclusions of the recent studies just mentioned) was the most advanced hospital of its time (fifth to tenth century) and it, most likely, served as the model for the development of the Islamic medieval hospital. The kind of medicine that was practiced at Jundisapur was greatly influenced by Indian medicine. However, the

[3] Jundisapur or Gundeshapur (Gund-ī Shāh Pūr, Gondeshapur, Jondishapoor, Jondishapur, etc.) means Army of Shapour. It was the intellectual center of the Sassanid empire and the home of the Academy of Jundisapur. (http://en.wikipedia.org/wiki/Gundeshapur – last visited March 30, 2012).

predominance or "molding" influence of Greek medicine during the development of the Jundisapur tradition cannot be denied. Jundisapur was, in fact, the foremost representative of Greek medicine of its time. In a similar way, the hospitals of medieval Islam can be thought of as the true forerunners of the modern hospital. This, because, in contrast to the Greek "Asclepions" which were – in fact – shrines, the Islamic hospitals were well-organized and specialized institutions of charity, and they constituted strongholds of scientific medicine. Turks seem to have played a prominent role in the development of Islamic hospitals. The fifth Islamic hospital was built by a Turkish general and minister (Fath ibn Khaqan ibn Gartuch), and the sixth (which was the first Islamic hospital supported by the waqf endowment) by Ahmed ibn Tulun. Out of the five earliest hospitals that had such endowment, the Turks built three or possibly four of them.

As for the Turkish achievements in the theory of Physiology and Medicine and the understanding of the function of the human body in general, it can be claimed that the ideas of Ibn al-Nafis of the thirteenth century influenced the discovery of the circulation of blood in Europe. This is so because Ibn al-Nafis first, ever, described the pulmonary circulation and his description was translated into Latin by 1547 or several years before the publication of the same discovery by Michael Servetus and Realdo Colombo. It is also claimed that the pulsilogium[4] whose invention is attributed to Sanctorius (1561–1636) was related to a similar instrument used by Sabunjuoglu Sherefuddin, a fifteenth century Turkish physician and surgeon from Amasya.

6.3 Who Was Spyridon Mavrogenis?

According to Theodore Blancard (2011),[5] the Mavrogeni family descends from the Venetian Morozini family. The Morozini family, itself, came from Byzantium, where they were known by the name Mavrogeni. Following the conquest of Byzantium by Sultan Mehmed the 2nd (Mehmed the Conqueror) the Mavrogeni family dispersed along the coasts of the Aegean and the Adriatic. A branch of the family ended in Peloponnese, another in Euboea and a third one in Crete. Francesco Morozini (1618–1694), a great Venetian admiral, landed in Peloponnese in June of 1685 and took it by storm. Two years later, in 1687, he occupied Athens.

The Mavrogeni name appears for first time in 1274, as Mario Mavrogeni or Morozini. According to Blancard, the relation between the two names is evidenced both by the phonetic and orthographic resemblance of the two names and by the use of the Lion of Saint Mark on the coat of arms of the Mavrogeni family in exactly the same form as it appears on the coat of arms of the Morozini family. Blancard maintains that when the Turks subjugated Peloponnese in 1715 two Venetian brothers, Stephanos and Petros Mavrogenis left Peloponnese along with their parents, their

[4] Levett and Agarwal (1979).

[5] Theodore Blancard "The House of Mavrogenis," Estia, Athens 2011.

Fig. 6.2 Spyridon Mavrogenis

brother (Nickolaos) and their sister to settle in the island of Paros. They were very wealthy, especially Petros, who was vice-consul of Austria and England in the Cycladic islands. Later, Stephanos moved to Istanbul, where, thanks to his education, he became dragoman in the Sultan's fleet. Petros Mavrogenis' son Nickolaos became governor of Moldavia and Wallachia (1786–1790).

Spyridon Mavrogenis (Fig. 6.2) was Petros Mavrogenis' great-grandson. According to his description on the front page of his book "Treatises on Physiology" he was "Professor Emeritus of All Medical Science, Member or president of many national and international scientific and literary societies, Chief Medical Officer and Personal Physician of His Majesty the Sultan and General Inspector of all the Medical Schools of Istanbul and all the Civilian and Military Hospitals in the Capital and the Provinces of the Empire… etc., etc., etc.[6]" It is indicative of Mavrogenis' influence in the final years of the Ottoman Empire that his son Alexandros Mavrogenis (Petros Mavrogenis' great great-grandson – 1848–1929) became Private Secretary to the Sultan, Ambassador to the United States, Governor of Samos Island in the Aegean and Senator.

6.4 The Structure of Spyridon Mavrogenis' Treatise

Following a brief introduction where Mavrogenis presents what he regards as the current thinking on the hard-problem (or "of "Life and Soul" and their interrelationships" as he names his treatise) and the preceding historical development of the subject, the text is split in two parts: The first deals with the definition of life, from

[6] "The office of Chief Medical Officer (*hakim-bashi*) was vested with following duties: medical care of Sultan & imperial family, palace staff, administration of all medical schools, physicians, ophthalmologists, & pharmacists." Quoted from: Zakaria Virk, Science and Technology in Ottoman Sultanate, http://www.alislam.org/egazette/articles/science-and-technology-in-ottoman-sultanate/ (last visited on March 30th, 2012).

Table 6.1 The structure of the Spyridon Mavrogenis Treatise "Of 'Life and Soul' and their interrelationships

Part A:	On Life
A.	The methodology for studying physical phenomena
	The limits of science
B.	The Laws of Nature
C.	On the teleology of natural phenomena
D.	On the nature of physical forces
E.	On mechanisms and organisms
F.	On the mechanical and dynamic behavior of the physical bodies
	On excitability and vitality
G.	The origin of beings
H.	Similarities and differences between organic and inorganic matter
I.	Applications
Part B:	On the Soul and its relationship to Life
A.	Mind (spirit) and matter – On the Soul and the Body – The Human Being
B.	A method by which we can understand the Soul; The distinction between perceptible and imaginary worlds
C.	Spiritualism, materialism, the system of identity. Physiology of the soul
D.	Principles of self-sense, common-sense, sensing the external world
E.	The brain and the nerves are the organs of sensation
F.	On touch
G.	On the gustatory sensation
H.	On olfaction
I.	On vision
J.	On hearing
K.	General comments on the nature of sensation and the brain as the ultimate sensory organ

a physics, chemistry and biology standpoint. The second treats the question of the soul and its relationship to life (Table 6.1).

For Mavrogenis there are two main principles that guide him in adopting this structure for his inquiry into the hard-problem:

1. "Everything that exists within nature, participates in the phenomenon of life, either explicitly or implicitly. Every thinking subject uses his mind to position himself as the center of things; pays attention to whatever phenomena assault his senses and his thinking process and, due to his human nature and innate drive, he undertakes to explore and explain everything that happens around him".
2. Man is bound by two main, innate and self imposed, constraints: "Know thyself" and "do not commit hubris"

Mavrogenis decides, therefore, to structure the first part of each main part of his text around the methodology that he deems best for studying the relevant phenomena. This is in line with the principle that was presented above or that the thinking human being is positioned at (and possibly serves as) the center of the explorable world. At the same time, he deliberately limits himself, in terms of his goals, when exploring the philosophical implications of his description of physical phenomena and of the phenomenon of life in particular; in fact, he deliberately describes three aspects (as he perceives it) of the hard-problem: the dipole spirit (mind) and matter, the dipole soul and body and, finally, the integration of these dipoles as to give rise to the Human Being.

6.5 The "Eastern" Perception of Western Thinking as It That Relates to the Hard-Problem

Mavrogenis' introduction to his treatise is of interest for a combination of two reasons: It adheres to a rather common (among his contemporary Greek scholars) rule that wants for the introduction to a treatise not to exceed about 1/6th of the total length of the treatise; It also attempts, for the shake of completeness, to present, in as much detail as possible, the current state-of-the-art and the historical development in the field of enquiry. This approach severely limits the breadth and scope of any short review. However, exactly for the same reasons, it underscores the author's personal biases and, potentially, it illuminates the depth of his understanding of the topic.

Mavrogenis' review starts with Empedocles, a Greek (fifth century BCE) pre-Socratic philosopher from Agrigento in Sicily (Magna Grecia). Empedocles is best known for having originated the notion that the world is composed of four elements; Earth, Water, Air and Fire. According to Empedocles these elements were mixed or separated, to give the different forms and phases of matter. The mixing was due to powers such as Love and Strife. In this sense, Empedoclean physics was influenced by Pythagorean thinking. According to Mavrogenis, Empedocles, in line with Pythagorean dogma, thought that human life sprang from an inner "sublime force and drive" that transformed matter into life. This "sublime force and drive" was related to the Fire element, as evidenced by the warmth that characterizes living bodies. At the same time, all four elements were influenced by external forces and were subject to eventual disorganization, dissipation and decay because of the friction these external forces imposed upon them. Mavrogenis points out that it is on the same basis that Hippocrates built his medical system of thought. He also points out that, based on the same premises, Plato arrives to the conclusion that the anatomical parts of a living body are held together, as a functional whole, by Fire and Air (spirit). Mavrogenis abruptly closes his review of antiquity by very briefly referring to Galen for whom he expresses his admiration because "he firmly held the scepter of medical science from the second to the sixteenth century CE" following the renaissance of Hippocratic thought that took place around the first century CE.

According to Mavrogenis, the recipient of Galen's "scepter" was Philippus Aureolus Theophrastus Bombastus von Hohenheim or Paracelsus. Mavrogenis compares Paracelsus to Luther inasmuch as Paracelsus "smashed" the Hippocratic scepter that he received from the hands of Galen" and replaced Earth, Water, Air and Fire with Sulfur, Salt and Mercury! This, according to Mavrogenis, Paracelsus only replaced the four, arbitrary in his view, Empedoclean elements with three others whose choice was just as arbitrary – not a terribly important improvement. Moreover, he points out, Paracelsus still had to postulate the existence of a vital force («Αρχαίον» or "Ancient") to be acting upon these "elements" for life to arise and he expresses his amazement that, for 2000 years since Empedocles, no essential progress had been made.

The next stop in Mavrogenis' account of the history of the "hard-problem" is the seventeenth century. He briefly refers to the thesis of Jan Baptist van Helmont "on spontaneous generation" and points out the reliance of his thinking on the existence of a vital force. This force, however, had a more tangible form as it was produced "from water under the influence of some kind of yeast". It presented with the properties of gases and was confined to the stomach from where, "like the captain of a ship governing the rudder," directed the whole body. Although not impressed by Jan Baptist van Helmont's thinking, Mavrogenis still believed that the origins of life should be sought through chemistry. Along these lines he presents the work of Franciscus Sylvius (Franz de le Boë) who regards the phenomenon of life as something akin to maintaining a solution at a particular pH and he points out that any disturbance of the pH of the solution (i.e., of the solution's acid–base balance) resembles the process of disease and (at the extremes) possibly death.

Following a brief reference to the scientific advances that, meanwhile, Galileo Galilei, Isaac Newton, Antoine Lavoisier and William Harvey had brought to science, Mavrogenis proceeds to present the theories that, in his view, that were popular in the eighteenth century. He proceeds to review the propositions of:

1. Giovanni Alfonso Borelli (life is a mechanical manifestation of friction between the solids and liquids that composed the human body),
2. Friedrich Hoffmann (the proper circulation of all humors through the nerves, the vascular system and the different excretory ducts; these conduits, when irritated by external forces alternatively contract and relax, giving thus rise to life),
3. Georg Ernst Stahl (phlogiston theory, animism as opposed to the materialism of and Friedrich Hoffmann).
4. Albrecht von Haller (any "remedy must be proved on a healthy body, without being mixed with anything foreign; and when its odour and flavour have been ascertained, a tiny dose of it should be given and attention paid to all the changes of state that take place, what the pulse is, what heat there is, what sort of breathing and what exertions there are. Then in relation to the form of the phenomena in a healthy person from those exposed to it, you should move on to trials on a sick body…")
5. Christoph Girtanner (oxygen as the basis of irritability) and the rest of the Lavoisier followers Mavrogenis completes his expose of eighteenth century thought by referring to the views of
6. John Brown (life as a unique expression of irritability) and
7. Erasmus Darwin (life as the result of releasing the "sensory energy" which is stored in the nerves and muscles of the body due to external stimuli).

It should be noted that both John Brown and Erasmus Darwin had been, one way or another, at odds with Christoph Girtanner.

Mavrogenis extends his review into his own century (nineteenth) with an exposition of Natural Philosophy, which he refers to as the "daughter" of Luitzi Galvani who was "nurtured" by Baruch Spinoza and finally "reared" by Friedrich Wilhelm Joseph Schelling (of "Naturphilosophie"). He concludes that the only advantage that Natural Philosophy offers, over its predecessors, is that it "promoted the scientific study of nature which", in a relatively short time, "brought about many discoveries". These, he proceeds to identify as:

1. The recognition that "force and matter" are, in effect, inseparable in nature
2. The realization that "all living organisms are related to each other", although he meticulously avoids to call the process by which this came about (despite clearly implying it), by the name it was already known in the West, i.e., "Natural Selection"!

Finally, he completes the introduction to his brief treatise by presenting a host of popular at his time definitions of what life is. Among them (the quotations are a liberal translation of the Mavrogenis' text), those of:

1. Immanuel Kant ("Life is the force which, by virtue of the internal momentum of its carrier, induces its carrier to act"),
2. Gottfried Reinhold Treviranus ("Life is the condition which is randomly promoted and preserved by external stimuli; when such stimulating phenomena are absent everything tends to return to equilibrium")
3. Carl Christian Erhard Schmidt ("Life is the force generated by active matter; it is directed by the active matter's own organizational laws")
4. Christoph Wilhelm Friedrich Hufeland ("The cause of life is a vital force, different from the physical, chemical or mental forces. It permeates all organic bodies although it presents itself in different ways, either as a productive force, or as a moving force or, sometimes, as the cause for sensation")
5. Friedrich Ludwig Kreysig ("The cause of life is a vital force which resides in the nervous system")
6. Friedrich Wilhelm Joseph Schelling ("Life is born out of the conflict between opposing physical forces and it is because of this conflict that an animal lives")
7. Karl Robert Edward von Hartmann (1876a, b) ("Life is the dynamic conflict of opposing principles taking place in each and every part of the body")

Mavrogenis concludes his introduction by stating: "I do not, a priori, propose any definition of life. However, before I embark upon a discussion on the nature of life, I prefer to search for the best method for understanding physical phenomena. I solemnly state my belief that the actions of the soul are a separate entity from those of life. Thus, because I do not think that our thoughts are secretions of our brain, I will examine both life and the soul separately, starting with the examination of life"

6.6 Spyridon Mavrogenis' Perception of Nature and His Position on the Applicability of the Natural Laws to Biology

In discussing the limits of science, Mavrogenis first makes reference to the historical transition from the synthetic methodologies of the past (whereby "Physicians and Philosophers" hoped to axiomatically analyze nature and progress from general principles to the particulars of a problem), to what was still recent for his era – the analytic methodologies. Not able (or, rather, willing) to take a position on which of the two methods is the best he expresses his preference for "the middle road". This, he summarizes in two simple statements:

1. "All things that can be perceived through our senses are not random or irregular but they obey immutable physical laws".
2. "Only the Supreme Being is self-caused and does not obey any laws"!

These statements he supplements by a comment as to the "purposefulness" of physical phenomena which can be liberally translated as follows: "By studying the phenomena taking place in the physical world as a chain of cause and effect, we slowly come to the realization that there is a purpose behind them. We will never be able to find the 'mysterious paths' by which God's wisdom brought about the apparent harmony in the Universe. Nor will we be able to understand the purpose of the different phenomena. For the purpose of the different phenomena does not exist before they take place but, rather, it becomes evident after they take place". Thus "not only the cause matches the effect but also both cause and effect match the purpose God has for the manifestation of the particular phenomena".

Following the statement of his position, Mavrogenis proceeds to discuss the nature of physical forces, paying particular attention at what the disciplines of Physiology and Biology of his time regarded as "vital forces". He contrasts the position of Gottfried Reinhold Treviranus about the indestructibility of matter which takes on different forms (depending on the external forces or influences) with that of Johann Heinrich Ferdinand Autenrieth who compared the nature of vital forces to other "weightless" forces or objects that the physics of his era was postulating to exist (for example the corpuscular nature of light). He concludes that the state of thinking in his era, as exemplified by these theories, does not differ from the fallacies of the ancients about the nature of disease. He suggests therefore not to pay so much attention to theory but to stay close to observation! Along this line of reasoning, he insists that nature is full of abstract laws, many of which he regarded as still unknown, but, yet, their effects could be described if one considered the human body as a machine. Thus, in his view, what was important was not the kind or form of the forces that exist in nature but the kind of organization that utilizes them.

Following the discussion on the nature of physical forces Mavrogenis proceeds to propose that the "human machine" is characterized by the property of "excitability" which he defines as “the property of animals to react to external stimuli (or forces)”. He makes it clear that the property of excitability is not an exclusive property of living things and he authoritatively states that excitability in itself does not define life. Then, after dwelling on a discussion about the origin of living beings and the similarities and differences between organic and inorganic matter he arrives at a series of conclusions or statements about the phenomenon of life:

1. Living organisms differ from non living matter only in design and complexity.
2. How this design and complexity evolve from the stage of the fertilized ovum to the complete organism is the major unresolved problem of Biology.
3. The purpose of life itself is, without doubt, the recognition by the living entity of the existence of its environment. This is coupled with the ability of living beings to independently and spontaneously move within their environment.
4. To attribute the phenomenon of life to the action of a vital force, not only do we delude ourselves but we, in effect, put great obstacles to any further study of the phenomenon of Life.
5. Life is characterized by the fact that, after death, each body part disintegrates at a rate proportional to that by which it was composed. Blood for example is replaced with great speed during life. It therefore it is the first tissue to disintegrate after death. In contrast bone takes a long time to build and similarly, it lasts the longest after death.
6. Life is not a force but a process
7. Life by itself represents an internal state of agitation and change. To be sustained it requires mechanisms by which the living organism responds to environmental challenges.
8. Life conforms to all physical laws.
9. In death, the organism disintegrates; however, the materials of which it is composed remain indestructible and are recycled into other organisms

6.7 Spyridon Mavrogenis’ Perception of the Soul and Its Relation to Life

In investigating questions relating to the existence of the Soul and its properties, Mavrogenis uses the same approach for building his arguments as he did when he dealt with questions that relate to issues of life. He insists that both analytic and synthetic methodologies should be employed, only this time he does not opt for the “middle road”. On questions about the Soul he totally detaches himself from materialism. He states: “We have seen how difficult it is to understand and explain the phenomena of life. However, we pointed out that within human beings something else, other than matter, also exists; something that is superior to the matter and which is not bound by physical laws; something non destructible, which obeys only to itself and which is fully cognizant of its existence. This we call mind or spirit. Thus the mind, the soul and the body, together, are the tripartite existence that constitutes every individual human being”.

Faced with the religious beliefs of his audience, which included some of the best clergy (well versed in Eastern Orthodox ecclesiastic dogma) and Rum (of Greek Orthodox persuasion) civil servants of the Ottoman Empire, Mavrogenis clearly takes a "politically correct" position. He states (in liberal translation): "What the mind consists of, we cannot understand. At the same time our acknowledgement that our mind is not subject to physical laws but to laws of a different nature than the natural laws, is not a scientific position; it, much rather, reflects our ignorance!"

He proceeds to speculate, without presenting any supporting evidence, on why anatomists, originally, started dissecting the brain by proposing that it was the generally accepted belief that cognition as well as morality have their seat in that organ. For Mavrogenis, it is this deeply rooted belief among mankind that "led to the realization that there is a world accessible to man through the mental processes and another world accessible to him through the sensory experience. He insists that these two notions, together, define humanity. He further attempts to strengthen his argument by underlining what to him was a fact, i.e., that all attempts to bridge the gap between the world accessible to man through mental processes and the world accessible to man through the sensory experience failed except those that relied on the assumption that a Supreme Being does exist! For Mavrogenis "The notion that the body does not differ from the mind or that the mind is just a higher manifestation of our body is plain wrong and materialism, as expressed by such ideas does not help solve the mystery of human existence but, rather, it makes it denser".

In line with Eastern Orthodox ecclesiastic dogma Mavrogenis states that "man himself is not spirit. However he is cognizant of the spirit in him," and he clearly rejects the proposition of Johannes Klencke that nature and mind are a single entity, insisting that there is a definite duality of nature and mind. In this respect he speculates that "we cannot understand our own self because we are our own self; the same way that our hand cannot feel itself". However, Mavrogenis proclaims that the duality that exists between nature or body on one hand and the mind, on the other, does not preclude their scientific exploration! For him, his position that "the body and the spirit are two separate entities," means that "the study of the body belongs to the realm of physics while that of the spirit belongs to the realm of metaphysics," elevating thus metaphysics to the rank of science rather than pure speculation.

6.8 Spyridon Mavrogenis' Conclusions

It is interesting, when Mavrogenis's dualistic approach to the soul and its relationship to the body is taken into account, his conclusion that the brain is the final storage area in the body for every notion (idea, impression, meaning), a "mirror where 'every notion' is reflected". Therefore, he continues, the brain "must be the principal organ of psychological activity". To support this he proceeds to lay down a series of clinical considerations that, in his opinion, strengthen his conclusions (presented in liberal translation within quotation marks):

1. “So long as no mental illness affects the brain, any body part may suffer damage or its capacity be diminished by disease without this affecting the subject’s mental capacity”.
2. “Many times we observe that despite severe mental trauma, the body (except for the brain) remains fully functional”. The fact that there are physical situations whereby the brain is incapacitated immediately after severe bodily damage, does not alter Mavrogenis’ stance on the matter; his explanation, for such cases, is that the body-soul connection is irreparably damaged or permanently severed.
3. “Organic brain damage does affect the mental condition of individuals. Strokes or skull fractures that result in” [increased intracranial pressure by the presence of] “hematomas alter the patient’s state of consciousness”.
4. “Anencephaly and other similar congenital abnormalities or disorders, are incompatible with life”.
5. “Once the organic cause that affects a brain is treated or removed, consciousness returns”
6. “The mental capacity” [mental worthiness] of living organisms is directly proportional to the degree of development of their brain. Likewise, the development of an individual’s mental capacity parallels that of his brain and its condition”
7. Once a nerve is damaged, so that its continuation with the rest of the nervous system is disrupted, the sensations it sub-serves diminish or are abolished”

How does Mavrogenis justify his insistence on the existence of soul in view of the, nearly perfect and pure, clinical reasoning that was just laid out? His answer introduces a new, rather sophist, twist in his argument: The reflections of different notions in the brain are meaningless as they do not contain meaning by themselves. For Mavrogenis, in order for the reflection of a notion to acquire meaning, the presence of the soul is necessary. He continues by saying: “The process of” [this] “transformation is still unknown to us. It occurs automatically, it is self originating, it is self governing and, as far as the process of understanding is concerned, it manifests itself as ‘attention’” . He borrows an analogy from Markus Herz, who likened the action of the soul as that of a mirror, whereby the mirror perfectly represents the image that impinges on it but cannot mirror itself! For Mavrogenis, therefore, the brain cannot be cognizant of itself unless a separate entity confers upon it the faculty of self cognition; self cognition is a property the soul confers and self cognition is responsible for directing attention!

References

Blancard T (2011) The house of Mavrogenis. Estia, Athens

von Hartmann E (1876a) Anfänge naturwissenschaftlicher Selbsterkenntniss (1873). In: Gesammelte Studien und Aufsätze gemeinverständlichen Inhalts. Duncker, Berlin, pp 445–459

von Hartmann E (1876b) Philosophie des Unbewußten, 7th edn. Duncker, Berlin

Levett J, Agarwal G (1979) The first man/machine interaction in medicine: the pulsilogium of Sanctorius. Med Instrum 13(1):61–63

Mavrogenis Spyridon. Treatises in physiology, ca. 1895–1900. “Astir” Publishers, Paris

Chapter 7
George Henry Lewes (1817–1878): Embodied Cognition, Vitalism, and the Evolution of Symbolic Perception

Elfed Huw Price

The history of science is the science itself: the history of the individual, the individual

Johann Wolfgang von Goethe, *Mineralogie und Geologie* ('Die Geschichte der Wissenschaft ist die Wissenschaft selbst, die Geschichte des Individuums das Individuum', Goethe 1833, 130)

Our world of Thought is a strange mixture of truth and fiction,—of Experience condensed in symbols, and of inferences deduced from symbols, and taken for reals; but the advance of Humanity tends more and more to enlarge the fund of truths, and to disclose the pitfalls on its path. The history of the race is but that of the individual "writ large."

George Henry Lewes, *Problems of Life and Mind* (Lewes 1875, 119)

7.1 Introduction

George Henry Lewes (pronounced Lewis) had a full and fascinating life, which has already been touched upon by biographers (including most recently Rosemary Ashton) yet he remains overlooked in most histories of neuroscience and philosophy. His life was unconventional from the outset. His father, John Lee Lewes, had two families and fathered two sets of offspring, one in Liverpool, and an illegitimate second brood in London (which was comprised of George Henry Lewes and his two older brothers). John Lee Lewes subsequently emigrated to Bermuda while Lewes was still very young. Lewes' mother told him his father had died (Lewes may have

E.H. Price (✉)
Independent Scholar, Oxford, UK
e-mail: huw.price@magdalen.oxon.org

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_7, © Springer Science+Business Media Dordrecht 2014

gone to his grave believing this to be true). His mother then married a retired captain from the East India Company army. Lewes had a rather peripatetic childhood and education in London, Jersey and France.

Lewes' career was comparably unorthodox. After an attempt at medical training, which he abandoned when he found he could not bear to see patients suffering, Lewes turned his hand to more literary pursuits and became a successful literary critic, journalist, and historical biographer as well as authoring several plays and novels. From 1836 he also harboured ambitions to write a synthesis of philosophy and physiology, although it was not until the 1850s that he began his physiological work in earnest. This coincided with the start of his cohabitation with Marian Evans, who was soon to become much better known as 'George Eliot'. They lived and worked together in a profound intellectual partnership until Lewes' death. Although Evans referred to herself as 'Mrs Lewes' their relationship was never legally recognized, Lewes already having an extant but failed marriage to Agnes Jervis from 1841. During the final decades of his life Lewes established his academic credentials through his work on the nature of the human body and mind. Scholars have noted the extent to which Lewes' physiological philosophy permeates Eliot's novels.[1] Yet, as substantive works meriting attention in their own right, Lewes' neurophilosophy of mind remains rather neglected.[2]

Lewes' late-bloom contribution to the sciences of life and mind was both respected and widely read; his *Sea-side Studies* (1858) tapped into contemporary interests in beachcombing; *Physiology of Common Life* (1859) was a success with medical students as well as a wider readership, and reached an international audience—the Russian luminary Ivan Pavlov credited this work with turning him to pursue physiology rather than a career in the church. Lewes' *Studies in Animal Life* was serialised in a new popularist periodical, *Cornhill Magazine*, in 1860. In 1862 Lewes, then in his mid-40s, turned his full attentions to 'the physiological mechanism of Feeling and Thought'.[3] This was an era in which the boundaries between physiology, psychology, philosophy and biology were in the process of being defined and delineated, and the very concept of clearly defined academic 'disciplines' was reaching acceptance.[4] The founding of *Nature* (1869), *Mind* (1876), and *Brain* (1878), now flagship journals for biology, philosophy, and neuroscience respectively, is indicative of this move towards disciplinary specialization. Lewes contributed to all three journals, a feat that reflects both the range and depth of his learning and the greater

[1] Shuttleworth 1984; A Ryan 2011; V Ryan 2012.

[2] There are a few notable exceptions, particularly Kaminsky 1952; Rylance 2000, 2004; and Tjoa 1977.

[3] Lewes 1874, vi.

[4] Stichweh 2001. Lewes was among the first to use the term 'discipline' in English in this context. Lewes 1874, p. 71. Lewes was not shy of introducing and promoting new terms and understandings. He has been credited with being the first Englishman to employ the terms 'psychodynamic' (Hearnshaw 1964), 'personality' (Rylance 2000) 'and 'stream of consciousness' (Holland 1986) in ways comparable to the present day use.

degree of overlap between these disciplines at the time.[5] Lewes held a central role in establishing the Physiological Society in 1876, another focal hub for disciplinary identity that remains extant today.

Lewes' theories about the mind received their most complete articulation in his *Problems of Life and Mind* (1874–1879), in which he endeavoured to set out a rational, scientific framework for approaching the metaphysics, physiology, psychology—and '*sociology*'—of the mind. Lewes viewed the human mind as equally the product of mankind's physical nature and nurture. This work, which ran across 5 volumes and well over 2,000 pages, was to be Lewes' swan-song. The final volumes of *Problems* lay unpublished at his death and were seen through to print by Evans, with the assistance of the renowned physiologists Michael Foster and James Sully. Evans also endowed a Cambridge Fellowship in physiology in his memory, the George Henry Lewes Trust, which would later support a young Charles Sherrington (amongst many others).[6] For reasons that remain obscure, Lewes' fame faded quickly after his death.[7] It may be that he was overshadowed by the dazzling literary success of his partner; or that his synthesis of neuroscience, philosophy, psychology, and sociology did not sit comfortably in an era of increasing disciplinary specialization. Or, it may be because many of his views ran against the accepted grain.

The aim of this chapter is to outline the conclusions Lewes reached in *Problems of Life and Mind*, focussing most closely on what might be viewed as his answer to the 'hard problem'. I will only offer a static snapshot of his beliefs he expressed during the last five years of this life. As Lewes readily acknowledged, in some aspects *Problems* marked a clear departure from views he had expressed in preceding works. Most notable was the shift away from his earlier rejection of metaphysical speculation as an endeavour doomed to failure, and therefore worthless, a stance Lewes had shared with, and possibly derived from, the arch-positivist French philosopher, Auguste Comte. In *Problems*, Lewes made a dramatic *volte-face* by placing topics which had traditionally been viewed as metaphysics at the heart of his research project.

For his time, Lewes was unusually aware of the embeddedness of beliefs and discoveries in their historical and social context.[8] Two of his most successful earlier works, his *Biographical History of Philosophy* (1845–1846), and his biography of Goethe (1855), addressed the place of significant philosophers within the broader history of philosophy. He would, I am sure, approve the rationale that lies behind the current volume. This chapter will not, however, offer more than a cursory account of the intellectual trajectory that eventually led Lewes to his most mature works, nor will it explore the key influences on his thought, although many of these are

[5] Lewes 1873, 1876a, b, c, 1877a, 1878.

[6] Tansey 1990, 1992.

[7] Ballantyne 1994.

[8] E.g., 'All germinal conceptions are the product of their age rather than of any individual mind', Lewes 1874, 84.

evident in *Problems*, and earlier works, in which he made no secret of his indebtedness to Spinoza and Comte.[9]

Problems of Life and Mind consists of 5 volumes split into three 'series'. The first series, 'The Foundations of a Creed' (2 volumes first published in 1874 and 1875) sets out the ground rules for his methodological approach. These volumes originally started as an introductory chapter to *Problems*, but ended up as a substantial series in their own right.[10] Here Lewes claimed to offer a new means to approach metaphysics by successfully adopting and applying scientific method to the areas susceptible to it, and leaving the rest alone: he christened the region of knowledge that currently lay beyond science, 'metempirics'.

The second series, a single volume on 'The Physical Basis of Mind', focuses on the physical (physiological) components of the phenomenon of mind, and the nature of life. This volume includes an essay on the nervous system, which sets out an understanding of nerve group function which Rick Rylance has described as 'strikingly modern' and reminiscent of Gerald Edelman's ground-breaking approach to the neuroscience of consciousness in the 1990s.[11] Similarly striking and modern, is Lewes' clear articulation of a hypothesis that the neuroglial cells—the 'other' (i.e. non-neural) type of cells that are found in the brain—play 'an essential part in all neural processes.' For most of the twentieth century, these cells (which are also known as glia) were generally viewed as little more than a kind of cerebral polystyrene, providing support for the nerves; yet in the last few decades they have increasingly been recognized as functionally integral components of the brain and even heralded as the key to unlocking the secrets of the mind.[12]

The third and final series of *Problems* consists of two volumes, both of which were published posthumously in 1879, and which address the other 'half' of mind, namely psychology and the interaction of man with society and history. A notice prefacing the first volume, which is by far the shortest volume of *Problems* at under 200 pages, states that this work was 'published separately in obedience to an implied wish of the Author, and has been printed from his manuscript with no other alterations than such as it is felt certain that he would have sanctioned'.[13] Entitled 'The Study of Psychology—its Object, Scope and Method', here Lewes aimed to build

[9] Lewes 1845, 1853. Lewes corresponded with Comte, and helped to bring his works to the attention of the English speaking public—including the first anglicized use of Comte's neologism, *'altruisme'*.

[10] Lewes 1874, viii.

[11] Rylance 2004; Edelman 1994.

[12] Fields 2009; Koob 2009. In the 1870s, at a time when his contemporaries were locked in debate over whether glia were neurons or connective tissue, Lewes disparaged this question as one that had been elevated to 'undue importance because it is supposed to carry with it physiological consequences which would deprive the neuroglia of active cooperation in neural processes, reducing it to the insignificant position of a mechanical support. I cannot but regard this as due to the mistaken tendency of analytical interpretation, which somewhat arbitrarily fastens on one element in a complex of elements, and assigns that one as the sole agent' Lewes 1877b, 246.

[13] Lewes 1879a, ix.

upon and beyond three traditional approaches to study of the mind: the philosophical tradition of Locke, Berkeley, Hume and Condillac; the physiological tradition of Cabanis and Gall, and the recent synthesis of the two offered by Bain and Spencer. The second and final posthumous volume—simply titled 'Problems of Life and Mind'—contained 'all the remaining manuscript' that Evans deemed to be in a state that Lewes would have sanctioned for publication.[14] This final volume of the series is split into three problems, or subsections: mind as a function of the organism; the sphere of sense and logic of feeling; and the sphere of intellect and logic of signs.

Given their central focus on the nature of life and mind, these five volumes in their entirety are of relevance to the 'hard problem'. Here I can only present selected excerpts and themes. The elements of greatest interest are not those that anticipate modern *established* theories, but rather those that remain provocative and challenging to this day. I will focus on three of these: firstly Lewes' understanding of the distributed (embodied) nature of sentience and consciousness; secondly, the grounds of his adoption of vitalism and rejection of mechanical materialism; and thirdly, his view of man as a uniquely symbolic species.

7.2 Mind, Body and Brain

In the last 20 years, there has been a surge—or resurgence—of interest in theories of embodied cognition: i.e., the view that other parts of the body in addition to, and beyond, the brain contribute to the processes of mind. Whilst this remains a fringe view, despite—or indeed, possibly *because*—of this, it continues to receive significant attention from philosophers. In 2009, Jesse Prinz observed that:

> Consciousness is trendy… Embodiment and situated cognition are also trendy. They mark a significant departure from orthodox theories, and are thus appealing to radicals and renegades. It's hardly surprising, then, that consciousness, embodiment and situated cognition have coalesced (see, e.g., Cotterill 1998; Hurley 1998; Mandik 1999; Noë 2005; O'Regan and Noë 2001; Thompson and Varela 2001). Both topics are exciting, and being exciting is an additive property. An embodied/situated theory of consciousness is the philosophical equivalent of a blockbuster.[15]

Even if most scholars remain sceptical, the role of embodiment in cognition and consciousness is a question that is now being taken seriously.[16]

Lewes' views, which encouraged an embodied view of the mind, touch on areas that are not only strikingly modern, but also strikingly radical, although, ironically, in his own day, they might have been viewed as outdated. During Lewes' lifetime Western conceptions of body and mind had undergone—and arguably were still in

[14] Lewes 1879b, v.

[15] Prinz 2009.

[16] E.g. See Wilson 2002; Clark 2009; Robbins and Aydede 2009.

the process of undergoing—a profound and dramatic transformation that would eventually result in a widespread emergence and acceptance of the notion that mind was a function of the brain. Yet this was a relatively recent (i.e. nineteenth century) development. In 1822, one promising London surgeon, William Lawrence, had been dismissed from his post at the lunatic asylum, Bridewell and Bethelem ('Bedlam') for daring to propagate the scandalous belief that mind was a function of the brain. His lectures on physiology—initially delivered at the Royal College of Surgeons in 1817, and published in 1819—were criticised by the clergy and condemned in the *Edinburgh Medical Review*, as 'calculated to lead the minds of his pupils into darkness worse than the darkness of the valley of death'.[17] The Anglican orthodox view, which had dominated the eighteenth century and beyond, advocated a strictly dualistic understanding of human rational self (ascribed to an immaterial and immortal soul), and a (more) distributed conception of the interaction between body and mind.[18] Through the eighteenth century, humoural theories inherited from Greek antiquity were increasingly abandoned in favour of ascribing individual differences in temperament to the quality and state of the nerves.[19] This was not, however, the same as ascribing mind to the brain. Brain-based doctrines, such as phrenology, were to remain anathema and were rejected by the British medical establishment up into the 1820s.[20]

However, within 50 years, the notion that mind was a brain function became commonplace. In *The Various Theories of the Relation of Mind and Brain Reviewed*, George Duncan (1869, 9) described three rival British physiological schools of thought: 'first, the Materialistic [sic], which asserts that the brain is the mind; the second, the popular school, asserts that the brain is the organ of the mind; and the third, which is the rising school, asserts that the mind is co-extensive with the nervous system, the brain being therefore only one of the organs of the mind.' Duncan identified Lewes as the leading exponent of the third school, which also happened to be Duncan's favoured stance.

In *Problems*, Lewes argued against the dominant physiological doctrine 'that the brain, and the brain only, is the source and seat of Sensibility'.[21] In 'The Foundations of a Creed', Lewes identified three assumptions which were generally taken to be axiomatic in physiological theories of brain function. Firstly, that the brain is the 'organ of the Mind'.[22] Secondly, that 'the mental diversities observable between

[17] As paraphrased in *The Times* 'Law Report', for the 25 and 27 March, 1822. For more on the Lawrence-Abernethy controversy, see Temkin 1963; Goodfield-Toulmin 1969; Butler 1993; and Price 2012.

[18] Yolton 1984; Price 2012.

[19] For theories of the nerves, see Rousseau 1991a, b.

[20] Shapin 1975; Cooter 1984.

[21] Lewes 1877b, 188.

[22] This tradition was pioneered at the start of the century by the London surgeon, William Lawrence and the followers of Franz Josef Gall, whose system of Organology (now better known as phrenology) held that not only intellect, but also emotions, mental illness, and will to be consubstantial with the brain. Price 2012.

men and animals, and between different races of men, are due solely to differences of cerebral mass'; and thirdly, that these differences in mental powers can be ascertained by examining the volume and weight of the brain. Lewes rejects all three axioms. 'The first, because to seek for an organ of the Mind is not less preposterous than to seek for an organ of the Life.' He moreover argued that, even if we were to accept the first premise, the second assumption would still be false in that it assigns 'the whole product to one of its factors' in that it overlooks the necessary and constructive role of culture (or to use his phrase, the 'Social Medium') in the formation of the mind. And against the third axiom, he points out that brain size and weight can tell us nothing about the cerebral differences that result from educational differences between races.[23]

In the third series of *Problems*, Lewes offers an additional critique of the cerebrocentric doctrine. He contended that physiologists had drawn false conclusions from the phenomenon of the spinal reflex, in particular the view that it was possible to divide nervous function neatly between 'conscious volitional' actions, on the one hand, and 'mechanical reflex' actions on the other.[24] According to Lewes, this division was erroneous, reached by starting from the mistaken premise that the brain was the exclusive somatic substrate of conscious experience. Because the reflex occurred without any need for cerebral intervention, and also without conscious experience it had been surmised that the reflex was purely mechanical. Lewes argued, to the contrary, that the phenomenon of *unconscious cognition*—of which Lewes was an early advocate—was sufficient to reveal the fallacy of this dichotomy. If mental processes were able to occur below the threshold of consciousness, then unless one also supposes these mental processes to be mechanical, unconsciousness cannot be equated with mechanism. 'We have no grounds for degrading any action of sentient mechanism from the psychical to the physical sphere, solely because it *might* pass unconsciously'.[25] Similarly, 'We must fix clearly in our minds that unconscious and insentient are not equivalent terms. … Unconsciousness is a sentient state, not the entire absence of Sentience we attribute to a machine.'[26] These views were allied to Lewes' conviction that sentience was a vital, and not just a physical mechanical process, which we will come to, below.

Throughout *Problems*, Lewes argued for a holistic, *embodied* conception of the mind: 'both physiologically and psychologically it is *we* who feel, and not any particular organ' and 'this *we* means the total sensibilities of the whole organism'.[27]

---

[23] Lewes 1874, 160–163. Gould 1981 presented a fascinating treatment of the history of craniometry and racism in *Mismeasure of Man*, although he does not consider the possibility raised here by Lewes—that not only was craniometry used to lend support to racism, but similarly racist prejudices conversely lent support to the adoption of the belief that that brain is the organ of mind.

[24] Lewes 1879a, 19–27. For a historical treatment of the reflex (although not Lewes' critique of it) see Clarke and Jacyna 1987, 101–156.

[25] Lewes 1879a, 23.

[26] Lewes 1879b, 150.

[27] Lewes 1879a, 20.

Whilst the nervous system played a central role in Lewes' system, he held that '*every neural phenomenon involves the whole Organism*'.[28]

7.3 Automata, Physicalism and Vitalism

In the first series of *Problems*, Lewes drew a radical distinction between living organisms—which he held to be purposive—and inorganic matter:

> by none of the laws hitherto detected in operation among inorganic phenomena it is possible to explain the biological laws of Nutrition, Evolution, Reproduction, and Decay. Should Molecular Dynamics one day be in a position to furnish such a deduction (which is probable), there would still remain the speciality of organic phenomena dependent on a speciality of concurrent causes, which would continue to separate biological from physical and chemical laws.[29]

Lewes thus rejected physicalism (in the sense of mechanical materialism) in favour of vitalism (with a lower case 'v'). This vitalism was not one that postulated the existence of immaterial spirits or supernatural elements.[30] Lewes stressed that organic matter differed from non-organic matter for reasons that he acknowledged to be as of yet unknown, although he intimated strongly that these reasons may be somehow associated with the distinctive chemical composition of organic matter.

> A biologist having ascertained that organic phenomena always require special combinations of oxygen, carbon, hydrogen, and nitrogen for their basis, and are never found where these are absent, accepts the ultimate fact of Vitality dependent on this combination. It is a fact no more explicable by reduction to some other fact, than why the ratio of 2/3 is the ratio of 4/6 or 8/12. The fact is so; is observed to be so; *why* it is so admits no further answer (for the present) than that whatever is is.[31]

Yet the 'speculative biologist' in Lewes' tale, unsatisfied with proclaiming a Socratic ignorance, instead posits the existence of an intangible 'Vital Principle' to explain the peculiar and unexplained phenomena of life. Whilst Lewes' classifies this vital principle as a fiction, he nonetheless holds it to be a useful concept, akin to the letter 'x' in algebra. It serves the purpose of representing an unknown, and thereby allowing the unknown to be brought within the scope of empirical enquiry.

In the second series, Lewes further elaborated on his view that living matter differed fundamentally from other matter:

> machines have fixed and calculated mechanisms; whereas organisms are variable and to a great extent incalculable mechanisms … a theory which reduces vital activities to purely physical processes is self-condemned. Not that we are to admit the agency of any

[28] Lewes 1874, 112: italics in original.

[29] Lewes 1874, 96.

[30] Lewes 1874, 110–111. Like his notion of sensibility, Lewes' materialist vitalism echoes theories expressed in the eighteenth century, most prominently by French physiologists such as Bordeu, La Mettrie and Bichat. See Haigh 1976 and Kaitaro 2008.

[31] Lewes 1874, 46.

> extra-organic principle, such as the [immaterialist] hypothesis of Vitalism assumes … but only the agency of an intra-organic principle… This assures us that an organism is a peculiar kind of mechanism, the processes of which are peculiar to it; and among those processes there is one which results in what we call Sensibility.[32]

Lewes traces the origins of the view of man as a machine (and with it, the modern formulation of the mind-body dichotomy) back to the division between two branches of natural philosophy: the mechanical objective view of nature developed by Galileo, Descartes, Newton and Laplace, and the subjective philosophical tradition of Locke, Berkeley, Hume and Kant. Following Comte, Lewes viewed the mechanisation of nature as a necessary but transitional stage in the progress of knowledge, allowing philosophers to move away from the primitive animist tendency to invest all objects with human qualities and motives. Yet this mechanization and de-animation of nature had been taken too far. Lewes noted that in his own time, psycho-physiologists had begun attempts to provide an objective account of feeling: 'to reduce Sensibility, in its subjective no less than in its objective aspect, to molecular movement'. Lewes considered that the mechanical doctrine was not appropriate or sufficient when it came to the analysis of life. Lewes instead aimed to provide a theory that stood between this physicalism and the opposing *Idealist* view which gave primacy to the realm of subjective phenomena, i.e., the view that 'things *are* just what they *appear*, since it is only in the relation of external reals to internal feelings that objects exist for us'. Lewes' vitalist philosophical physiology steered a middle ground between these two poles, through what he called 'reasoned realism'.[33]

One of the key problems facing the mechanical materialist worldview, in the nineteenth and twenty first centuries alike, is that it invites the problematic depiction of man as an automaton: if the cosmos is susceptible to a full explanation in terms of physical laws of cause and effect, there is no room, and no need, for free will and subjective experience. The spectre of epiphenomenalism, which was famously and controversially raised by Thomas Henry Huxley in 1874, continues to present a substantive challenge to materialist doctrines of mind. In a more recent incarnation, David Chalmers has switched metaphors, and replaced automata with 'zombies', but the essence of the debate remains the same. If man is a physical machine then what need is there for conscious awareness? The problem of automata was a heated issue in the years that witnessed the publication of *Problems* (i.e. 1874–1879). This brief period saw an efflorescence of writings on automata question, from the leading intellectuals of the age, including Huxley (1874), John Tyndall (1874), W. K. Clifford (1874), W. B. Carpenter (1874, 1875), Douglas Spalding (1877), William James (1879) and Henry Maudsley (1880). I will not delve into this debate here, but the reader will be well served for context through the existing secondary literature and the debates of this era continue to draw scholarly attention.[34]

32 Lewes 1877b, 324.

33 Lewes 1877b, 307–310. I have adopted the phrase 'psycho-physiologists' from Danziger 1982.

34 Daston 1978; Gray 1968; Greenwood 2010; Smith 2007.

Lewes' answer to the problem of automata was to reject physicalism and to advocate a difference of kind between the laws of organic matter and the laws of the inorganic cosmos.

For Lewes, there was a key concept missing from existing physicalist accounts of the mind. Lewes attempted to remedy this by introducing a revised notion of 'sensibility'. Sensibility was a prominent literary trope of the eighteenth century, closely associated with medical theories of the nerves, with roots in late seventeenth century sensationalist (Lockean) philosophy. In Lewes' time, as noted above, physiologists had started to try to reduce sensibility to physical phenomena. In *Problems*, Lewes reclaimed the concept of sensibility for nineteenth century science, although he recognized that it would not be easy for his contemporaries to assimilate or accept his revised vitalist interpretation.[35]

> Sensibility is a factor which raises the phenomena into another order. To overlook its presence is fatal to any explanation of the organic mechanism. Yet it is overlooked by those who tell us that when an impression on a nerve is conveyed to the brain, and is thence reflected on the limbs—as when the retina of a wolf is stimulated by the image of a sheep, and the spring of the wolf upon the sheep follows as a 'purely mechanical consequence—the whole process has from first to last been physical.' Unless the term *physical* is here used to designate the *objective sequence* as contemplated by an onlooker, who likens to the process to the sequence observable in a machine, I should say that from first to last the process has been *not* physical, but *vital*, involving among its essential conditions the peculiarly vital factor named Sensibility.[36]

Lewes held sensibility to be distinct from consciousness—which he described as 'a Mode of Sensibility'.

7.4 The 'Social Medium' and Symbolic Thought

In the introduction to the very first volume of *Problems*, Lewes reflects on the path he had followed up to that point: he remarks that he had passed from his study of the physiology of the nervous system (in the 1850s) to the study of animal psychology in the belief that this might reveal more about the nature of the human mind. Lewes' research on animals was unapologetically hands-on, practical and intensive, including dissections, and vivisection experiments on the effects of removing the brains of frogs.[37] In an 1859 letter to Lewes' son, Evans wrote, 'I wish you could have seen today, as I did, the delicate spinal cord of a dragon-fly—like a tiny thread with tiny beads on it—which your father had just dissected! He is so wonderfully clever now at the dissection of these delicate things, and has attained this cleverness entirely by devoted practice during the last three years'.

[35] Lewes 1877b, 188.

[36] Lewes 1877b, 324–325.

[37] Menke 2000.

Yet by the time Lewes came to write *Problems* he had abandoned studying animal cognition as a means of understanding man. He had, he noted, originally been 'misled by the plausible supposition that the complex phenomena in Man might be better interpreted by approaching them through the simpler phenomena in Animals,' but he had subsequently come to realize that the converse was true: 'to understand the mental condition of Animals we must first gain a clear vision of the fundamental processes in Man'.[38] The cardinal problem was the evident chasm between the abilities of man and animals, as a result of mankind's ability to symbolize and to use language. Lewes argued that animals were 'incapable of one supremely important mode of thinking—the formation of conceptions, and the combinations of series of feelings by means of verbal symbols. This, to which the name of Ideation may specially be given, is the distinguishing attribute of man'.[39] These symbols were mediated by language and embedded in society. Moreover, any attempt to describe animal behaviour inevitably led Lewes to employ descriptive language that was subjective and hence anthropomorphic, e.g. describing bees as being 'angry'.[40]

> Language by its generalisation enables us to construct *objects*—in the philosophical sense of the term—by separately naming, and thus giving separate ideal existence to, those feelings of a group which are invariable and predominant, as distinguished from the feelings which are variable and accidental. That is to say, the separation of subject and predicates, substance and attributes, object and qualities, thing and relations. No such separation can take place in the mind of the animal, nor in that of the infant… To animal and child, as we have said, subject and predicate are one.[41]

This was a bold statement. Contrast with the view expressed by one of Lewes' contemporaries, the Italian philosopher Tito Vignoli, 'The primitive and constant act of *all animals*, including man, when external or internal sensation has opened to them the immense field of nature, is that of entifying the object of sensation, or, in a word, all phenomena' (emphasis added).[42] Whilst Lewes shared Vignoli's view that entification (Lewes, more appropriately for his theory, referred to the process as 'personification') permeated cognition, Vignoli's attribution of this capacity to animals was (according to Lewes' reading) an anthropocentrism. Lewes' interpretation clearly lends itself to the view that the ability to abstract and personify evolved and came into being with man.[43]

While man and animal were both able to feel and experience, man alone was subject to, and immersed in the 'Social Medium—the collective accumulations of centuries, condensed in knowledge, beliefs, prejudices, institutions, and tendencies'. The social medium held a dual significance for Lewes. Firstly knowledge was a

38 Lewes 1874, v.

39 Lewes 1879b, 484.

40 Lewes 1879a, 119–121. Lewes was quite possibly the first to use 'anthropomorphic' in this modern sense, in his *Sea-side Studies* of 1858: c.f., Ashton 1991, 187.

41 Lewes 1879b, 489–490.

42 Vignoli 1885, 135–154.

43 Lewes was an early supporter of Darwin's theory of natural selection – cf. Ashton 1991, 243–245.

social product that was in a state of permanent and developing flux and progress. Secondly, cognition itself was dependent upon the social medium. Lewes stressed that, ‘Man is not simply an Animal Organism, he is also an unit in a Social Organism. He leads an individual life, which is also part of a collective life. …. Human Psychology, therefore, the science of psychical phenomena, has to seek its data in Biology, and in Sociology’.[44]

For Lewes, language was an essential component of perception. Whilst an animal (assuming it was not colour blind) would *feel* a difference between viewing different colours, it would not be able to *perceive* it as man does. The ‘logic of signs’ was surmounted on the ‘logic of feeling’, facilitated by symbols, and inextricably enmeshed within culture. ‘The possibility of this perception is due to Language; and Language exists only as a social function’.[45]

Lewes was not alone in defending the significance of the social medium—Herbert Spencer similarly allowed society a key role. However, Lewes went further than his colleague, allowing the social medium fuller reign and expression. In an otherwise critical review of Lewes’ posthumous volumes of *Problems*, in 1881, Carveth Read (Professor of Philosophy at UCL) conceded that: ‘it will perhaps be just to credit Lewes with the doctrine of the dependence of the Human Mind upon the Social Medium as his signal and crowning discovery … this truth is of the first rank in importance; once explained it is unquestionable; and Lewes has the sort of claim to have originated it that Bacon has to be considered the discoverer of Inductive Method: he was certainly not the first to apply it … but … was among the first to give it a deliberate and explicit enunciation and to predict its power and fruitfulness’.[46]

7.5 The Twenty First Century Relevance of Lewes’ *Problems*

Histories of nineteenth-century neuroscience have tended to focus on the pioneers of brain-based theories of mind, men such as Thomas Laycock, John Hughlings Jackson, and David Ferrier. For a brief period Lewes stood shoulder to shoulder with such men, but within a few years after his death his dissenting voice had been forgotten, or dismissed. Duncan’s assessment of Lewes’ system as ‘the rising school’ was to prove premature and optimistic. Lewes was swimming against the tide. By 1890 William James could assert, without apology, ‘that the brain is the one immediate bodily condition of the mental operations is indeed so universally admitted nowadays that I need spend no more time in illustrating it, but will simply postulate it and pass on’.[47] Yet, today, adherents of the doctrine of embodied cognition, such as those cited by Prinz in the passage quoted earlier in this chapter,

[44] Lewes 1874, 109.
[45] Lewes 1874, 124.
[46] Read 1881, 498.
[47] James 1890, 5.

are challenging this view. Lewes' system may offer a fresh and alternative perspective on this problem for a new generation of scholars.

Similarly, Lewes' critique of mechanical materialism remains as relevant now as it was in 1879. In the first volume of *Problems*, he observed that:

> Vitality and Sensibility may be said to rest on seriated Change. If the changes were simply movement, propulsive or combined, physical or chemical, they would not present the phenomena of Life or of Consciousness. These changes must be serial, and what we term organised, to present the phenomena of Evolution. That Life is Change, and that Consciousness is Change, has always been affirmed. We have only to add that the changes are serial, and convergent through a consensus determined by essential community of structure, and we have characterised the specialty of organic chance, demarcated Life and Mind from all inorganic change.[48]

With due caution in opening up a weighty topic, Lewes has hit upon a deep philosophical problem about the nature of life. The microscopic laws of physics are time symmetric (they contain no intrinsic 'arrow of time').[49] The march of science, including the discovery of DNA and the double helix, has not yet been able to explain the fact that living organisms develop serially, i.e. in a unilinear direction in time. Consciousness itself (experience) is so universally directional (time always goes forwards) that we rarely stop to question why this is so. Yet philosophy and science currently offer no compelling answer to why time does go forwards, rather than backwards, or bidirectionally. In other words, Lewes' point here may be valid—the serial nature of life does require explanation and it does distinguish it from inorganic matter. Moreover, Lewes' attitude towards mind and consciousness—as relying on a property of some yet to be fully understood element of living matter—resonates with much more recent exciting—and rather fringe and radical—claims made by Roger Penrose and Stuart Hameroff, which are addressed later in this volume. Lewes was writing not simply on the 'Problems of Mind' but the 'Problems of *Life* and Mind'.

Finally, Lewes' comments on symbolic cognition remain relevant to debates that straddle disciplines. The extent to which Lewes recognizes the grounding of science and knowledge in society and culture invites comparison with methodological approaches adopted 60 years later by Ludwig Fleck and championed by the proponents of Sociology of Scientific Knowledge a full century after Lewes.[50] Lewes' writings on the undeniable importance of culture (the 'Social Medium') and particularly his emphasis on the central importance of symbolic thought in human cognition chime with recent writings of Terrence Deacon (1997, 2011) and Merlin Donald (1991, 2001). One of the more prominent and controversial themes in *Problems* is Lewes' interpretation of the role and evolutionary origin of symbols,

---

48 Lewes 1874, 120–121.

49 See Huw Price (no relation) 1996 and Zeh 2007. Contrary to popular belief, temporal symmetry can also be argued for second law of thermodynamics, which is often taken to be archetypal directional law—see Jos Uffink 2001. Temporal asymmetry has been reported at the subatomic level in k-mesons (Angelopoulos et al. 1998) and b-mesons (Lees et al. 2012) yet it remains to be seen how or whether this relates to the macroscopic/subjective arrow of time.

50 Fleck 1979.

as a capacity unique to man. This contention remains heated today. Some scholars, such as Deacon and Donald view symbols as the basis of the intellectual leap into language and culture; others, such as the linguist Derek Bickerton, hold that the ability to symbolize is shared by our nearest living primate relatives, and even unrelated species such as parrots.[51] Lewes' work is both overlooked and relevant to these debates in equal measure. His system held that society provided the content and form for much of our mental life and had the evolution of linguistic (and symbolic) capacities of man at its heart.

7.6 Lewes and the Hard Problem

In its most succinct formulation, the hard problem asks: why are physiological processes 'accompanied by an experienced inner life'?[52]

In Lewes' framework, this question can and must be split into two.

Firstly 'why does sensibility (conscious or unconscious) occur?' And secondly, 'why does conscious experience accompany some forms of sensibility and not others?' From within Lewes' system, the former question is akin to asking why anything exists at all. For Lewes, feeling is fundamental. 'The starting-point is always Feeling, and Feeling is the final goal and test. Knowledge begins with indefinite Feeling, which is gradually rendered more and more definite as the chaos is condensed into objects, effected through a rudimentary analysis determined by the fundamental Signatures (Qualities) of Feeling, namely, Tension, Intension, Extension, Duration, Likeness, Unlikeness.'[53] Lewes recasts science itself, not as the objective ideal, devoid of personal experience but rather the contrary. It is built from experience. 'Experience is the registration of feelings and the relations of their correlative objects. Science is the explanation of these feelings, the analyses of these objects into their components and constituents'.[54]

As I have outlined, Lewes' contribution to this problem (I hesitate to call it an 'answer') is to (i) to attribute paraphysical ('vital') powers of organization and historicity to living things, (ii) to suppose that, via these vital powers, the nervous system facilitates *sensibility* or feeling, an irreducible concept that has no direct parallel in Chalmers' lexicon. (iii) This Sensibility need not entail conscious

[51] See Christiansen and Kirby 2003, especially the chapters by Bickerton (who refers to Irene Pepperberg's famous research on the capacities of parrots) and Davidson. Study of the evolutionary origins of language, *as an academic discipline*, is a relatively young field. However, as a subject of study, it is older than its advocates appear to realize. In recent works on the subject it has repeatedly been asserted that speculation on the evolutionary origins of language was brought to a halt (for 100 years) in 1866 by a proclamation of the Parisian Linguistic Society (e.g. Deacon 1997; Christiansen and Kirby 2003). Gregory Radick's historical account of research on primate language very effectively (if unintentionally) overturns this assertion: Radick 2007.

[52] Chalmers 1996, vii. See also Chalmers 1995.

[53] Lewes 1874, 101.

[54] Lewes 1874, 100.

awareness at all, but provides the basis for the logic of signs—the capacity to create abstract concepts, the building blocks of thought.

Lewes' approach offers an alternative perspective on Benjamin Libet's influential studies from the 1980s, which revealed reliable physiological indicators of willed action which precede and anticipate the subsequent subjective conscious experience of initiating an action. In other words, a physiological unconscious 'decision' is made to act before we are consciously aware of this. These experiments highlight the difficulty (under the traditional physicalist model) of reconciling mechanical materialism and the convincing subjective experience of having choice and responsibility in our actions.[55]

Lewes' answer is to reject the view that organic life is mechanistic in the same sense as inorganic matter. 'Owing to the popular misconception of the term Mechanism when applied to organisms, there is the notion that if our actions are mechanically determined they must have the fixity of invariableness observed in machinery'. Yet 'Consciousness assures us that our actions are not thus invariable.'[56] Lewes' system may offer a (retrospectively) novel means to approach this problem, as it would deny that the anteceding physiological indicators are any more mechanistic than consciousness itself. Both are expressions of sensibility or sentience. Free will does not necessarily have to begin with conscious awareness of it. Lewes held that his contemporaries had failed to distinguish between 'Sentience, the activity of the neuro-muscular system, and Consciousness (in the special sense of Reflection), the particular Mode of Sentience.' Or in other words, consciousness and sentience had been falsely conflated, just as sentience had wrongly been associated with the brain alone.

7.7 Conclusion

Writing in the 1870s, in the last years of his life, Lewes gave the world a monist, materialist, vitalist, social constructionist, evolutionary theory of consciousness.

Lewes advocated a monist doctrine that mind and matter are one and the same substance. He has been credited with pioneering the philosophical stance that has subsequently been dubbed 'neutral' or 'dual aspect' monism—body and mind being different aspects of the same reality.[57]

> A neural process may be regarded either as a physiological process of molecular changes in the nervous system, or as a psychological process of sentient change; but in reality it is always one process in a complex of related changes; *its physiological or its psychological character necessarily results from its relation to those changes which precede and those which accompany it*. We isolate it by an artifice. When isolated, it is no longer a state of Feeling, no longer a fact of Consciousness, except by the *implied* relations from which we

[55] Libet et al. 1983, 2004; Tallis 2011.

[56] Lewes 1879b, 111.

[57] Wozniak 1992, 12.

> have detached it. When restored thus to its real position as one process in the complex of vital processes, it ceases to be a physical fact, it becomes a vital fact; and this vital fact has two aspects—one physiological, in which it is a state of the organ, the other psychological, in which it is a state of Feeling.[58]

In the opening chapter of 'The Foundations of a Creed' Lewes introduced his work as tailored towards providing a path to certainty that had been left unfulfilled by the competing religious sects of the day. This new creed was to be justified through application of the scientific method to all aspects of the mind. Lewes was no atheist—he held science and religion to be ultimately reconcilable. He viewed his own work as an attempt to establish a 'Religion founded on Science'.[59]

Lewes was a materialist in that he recognized and shunned immaterial principles and substances as a product of the human Ideation. Lewes was a vitalist in that he observed and posited an intrinsic and radical difference between organic life and inorganic matter. Lewes might therefore be viewed as an eliminative vitalist monist–his 'resolution' of the hard problem of consciousness was one that would require a new physiological terminology. Yet, rather than ousting folk psychology terms, Lewes saw a need for physiology to adapt to accommodate common sense experience and feeling ('sensibility').

Lewes describes his stance as 'reasoned realism'. This is distinguished both from the Idealist view that to exist is to be perceived (Berkeley's *esse est percipi*) and the Realist view that objects have an objective existence aside from perception. Lewes argued that the only things we can discuss are those that we can perceive. As for that which lies beyond, we must remain silent.

Although Lewes raises as many questions as he 'solves', his approach may offer ways to transmute the insoluble hard problem into others which are potentially capable of answer, namely: what role (if any) does the body (embodiment) play in the formation of conscious thought? What is at the root of the difference between living and nonliving matter? Is organic matter 'incalculable'? And if so, why? Is symbolic thought (or some element of it) unique to man, and if so, how did it evolve? Historical research is an engagement with the past. It is no less an engagement with the present. Lewes' work bridges concerns that are as relevant today, if not more so, than they were in 1870s.

References

Angelopoulos A, Angelos A, Apostolakis E, Aslanides E, Backenstoss G, Bargassa P, Behnke O, Benelli A et al (1998) First direct observation of time-reversal non-invariance in the neutral-kaon system. Phys Lett B 444(1):43–51

Ashton R (1991) G. H. Lewes: a life. Oxford University Press, Oxford

[58] Lewes 1879b, 149.

[59] Lewes 1874, 3.

Ballantyne PF (1994) G. H. Lewes: early reviews to mid-20th century obscurity. http://www.comnet.ca/~pballan/LEWES.htm. Viewed and archived on 28 Aug 2010 using WebCite® at http://www.webcitation.org/5sK1Jzaf9
Butler M (ed) (1993) Mary Shelley's Frankenstein, or, the modern Prometheus: the 1818 text. Oxford University Press, Oxford
Carpenter WB (1874) Principles of mental physiology. H. S. King & co., London
Carpenter WB (1875) The doctrine of human automatism. Sunday Lecture Society, London
Chalmers D (1995) Facing up to the problem of consciousness. J Conscious Stud 2:200–219
Chalmers D (1996) The conscious mind: in search of a fundamental theory. Oxford University Press, Oxford
Christiansen MH, Kirby S (eds) (2003) Language evolution. Oxford University Press, Oxford
Clark A (2009) Spreading the joy? Why the machinery of consciousness is probably. Still in the head. Mind 118:963–993
Clarke E, Jacyna LS (1987) Nineteenth-century origins of neuroscientific concepts. University of California Press, Berkeley
Clifford WK (1874) Body and mind. Reprinted in: Stephen L, Pollock F (eds) (1886) Lectures and essays. Macmillan, London, pp 244–273
Cooter R (1984) The cultural meaning of popular science: phrenology and the organisation of consent in nineteenth-century Britain. Cambridge University Press, Cambridge
Cotterill R (1998) Enchanted looms: conscious networks in brains and computers. Cambridge University Press, Cambridge
Danziger K (1982) Mid-nineteenth-century British psycho-physiology: a neglected chapter in the history of psychology. In: Woodward WR, Ash MG (eds) The problematic science: psychology in nineteenth-century thought. Praeger, New York, pp 119–146
Daston LJ (1978) British responses to psycho-physiology, 1860–1900. Isis 69:192–208
Deacon T (1997) The symbolic species: the co-evolution of language and the brain. W.W. Norton, New York
Deacon T (2011) Incomplete nature: how mind emerged from matter. W.W. Norton, New York
Donald M (1991) Origins of the modern mind: three stages in the evolution of culture and cognition. Harvard University Press, Cambridge, MA
Donald M (2001) A mind so rare: the evolution of human consciousness. Norton, New York
Douglas Fields R (2009) The other brain: from dementia to schizophrenia, how new discoveries about the brain are revolutionizing medicine and science. Simon & Schuster, New York
Duncan G (1869) The various theories of the relation of mind and brain reviewed. Trübner & co., London
Edelman G (1994) Bright air, brilliant fire: on the matter of the mind. Penguin, London
Fleck L (1979) The genesis and development of a scientific fact. Edited by Trenn TJ, Merton RK; Foreword by Kuhn T. University of Chicago Press, Chicago
Goethe JF (1833) Goethe's Nachgelassene Werke. Cotta, Stuttgart
Goodfield-Toulmin J (1969) Some aspects of English physiology: 1780–1840. J Hist Biol 2:283–320
Gould SJ (1981) The mismeasure of man. Norton, New York
Gray PH (1968) Prerequisite to an analysis of behaviorism: the conscious automaton theory from Spalding to William James. J Hist Behav Sci 4:365–376
Greenwood J (2010) Whistles, bells, and cogs in machines: Thomas Huxley and epiphenomenalism. J Hist Behav Sci 46:276–299
Haigh EL (1976) Vitalism, the soul, and sensibility: the physiology of Théophile Bordeu. J Hist Med All Sci 31:30–41
Hearnshaw LSA (1964) Short history of British psychology, 1849–1940. Barnes & Noble, New York
Holland JG (1986) George Henry Lewes and 'stream of consciousness': the first use of the term in English. South Atl Rev 51:31–39
Hurley S (1998) Consciousness in action. Harvard University Press, Cambridge, MA

Huxley TH (1874) On the hypothesis that animals are automata, and its history. Fortnightly Review 22:555–580. Reprinted in Collected essays: Vol I, Method and results. Macmillan, London, pp 195–250, 1893
James W (1879) Are we Automata? Mind 13:1–22
James W (1890) The principles of psychology. Macmillan, London
Kaitaro T (2008) Can matter mark the hours? Eighteenth-century vitalist materialism and functional properties. Sci Context 21:581–592
Kaminsky J (1952) The empirical metaphysics of George Henry Lewes. J Hist Idea 13:314–332
Koob A (2009) The root of thought: unlocking glia—the brain cell that will help us sharpen our wits, heal injury, and treat brain disease. FT Press, Upper Saddle River
Lawrence W (1819) Lectures on physiology, zoology and the natural history of man. Longman, London
Lees JP, Poireau V, Tisserand V, Garra Tico J, Grauges E, Palano A, Eigen G et al (2012) Observation of time-reversal violation in the B^0 meson system. Phys Rev Lett 109(21):211801
Lewes GH (1845–1846) A biographical history of philosophy, 2 vols. Charles Knight & Co., London
Lewes GH (1853) Comte's philosophy of the sciences: being an exposition of the principles of the Cours de Philosophie positive of Auguste Comte. HG Bohn, London
Lewes GH (1855) The life and works of Goethe. D. Nutt, London
Lewes GH (1858) Sea-side studies at Ilfracombe, Tenby, the Scilly isles, & Jersey. W. Blackwood, Edinburgh
Lewes GH (1859–1860) Physiology of common life, 2 vols. W. Blackwood, Edinburgh
Lewes GH (1873) Sensation in the spinal cord. Nature 9:83–84
Lewes GH (1874) Problems of life and mind first series: the foundations of a creed, vol I. Trübner & co., London
Lewes GH (1875) Problems of life and mind first series: the foundations of a creed, vol II. Trübner & co., London
Lewes GH (1876a) Review of F. Lussana and A. Lemoigite, Fisiologia dei Oentri Nervosi Encefalici. Mind 1:122–125
Lewes GH (1876b) Review of Ferrier's functions of the brain. Nature 12, Nov 23 & Nov 30
Lewes GH (1876c) The uniformity of nature. Mind 2:283–284
Lewes GH (1877a) Consciousness and unconsciousness. Mind 6:157–167
Lewes GH (1877b) Problems of life and mind second series: the physical basis of mind. Trübner & co., London
Lewes GH (1878) Motor feelings and the muscular sense. Brain 1:14–28
Lewes GH (1879a) Problems of life and mind: third series: the study of psychology. Its object, scope, and method. Houghton, Osgood and Co., Boston
Lewes GH (1879b) Problems of life and mind: mind as a function of the organism. The sphere of sense and logic of feeling. The sphere of intellect and logic of signs. Trübner & co., London
Libet B (2004) Mind time: the temporal factor in consciousness. Harvard University Press, Cambridge, MA
Libet B, Gleason CA, Wright EW, Pearl DK (1983) Time of conscious intention to act in relation to onset of cerebral activity readiness-potential. The unconscious initiation of a freely voluntary act. Brain 106:623–642
Mandik P (1999) Qualia, space, and control. Philos Psychol 12:47–60
Maudsley H (1880) The physical basis of will. Sunday Lecture Society, London
Menke R (2000) Fiction as vivisection: G. H. Lewes and George Eliot. ELH 67:617–653
Noë A (2005) Action in perception. MIT Press, Cambridge, MA
O'Regan JK, Noë A (2001) A sensorimotor account of vision and visual consciousness. Behav Brain Sci 24:939–1031
Price H (1996) Time's arrow and Archimedes' point: new directions for the physics of time. Oxford University Press, Oxford
Price EH (2012) Do brains think? Comparative anatomy and the end of the Great Chain of Being in 19th-century Britain. Hist Hum Sci 25(3):32–50

Prinz J (2009) Is consciousness embodied? In: Robbins P, Aydede M (eds) Cambridge handbook of situated cognition. Cambridge University Press, Cambridge
Radick G (2007) The simian tongue: the long debate about animal language. University of Chicago Press, Chicago
Read C (1881) G. H. Lewes's Posthumous volumes. Mind 6:483–498
Robbins P, Aydede M (eds) (2009) Cambridge handbook of situated cognition. Cambridge University Press, Cambridge
Rousseau GS (1991a) Nerves, spirits and fibres: towards an anthropology of sensibility. In: Rousseau GS (ed) Enlightenment crossings: pre- and post-modern discourses: anthropological. Manchester University Press, Manchester, pp 122–141
Rousseau GS (1991b) Towards a semiotics of the nerve. In: Burke P, Porter R (eds) Language, self and society: the social history of language. Polity Press, Cambridge, pp 213–275
Ryan AE (2011) Victorian fiction and the psychology of self-control, 1855–1885. Doctor of Philosophy, Case Western Reserve University, English, 2011
Ryan VL (2012) Thinking without thinking in the Victorian novel. Johns Hopkins University Press, Baltimore
Rylance R (2000) Victorian psychology and British culture 1850–1880. Oxford University Press, Oxford
Rylance R (2004) Convex and concave: conceptual boundaries in psychology, now and then but mainly then. Vic Lit Cult 32:449–462
Shapin S (1975) Phrenological knowledge and the social structure of early nineteenth-century Edinburgh. Ann Sci 32:219–243
Shuttleworth S (1984) George Eliot and nineteenth century science. The make-believe of a beginning. Cambridge University Press, Cambridge
Smith CUM (2007) William Carpenter and the Debate on Human Automatism in mid-Victorian England. In: Proceedings of the international society for the history and social studies of biology, 2007. University of Exeter, Exeter
Spalding D (1877) Review of problems of life and mind. Nature 10:261–263, August 2
Stichweh R (2001) History of scientific disciplines. In: Smelser NJ, Baltes PB (eds) International encyclopedia of the social and behavioral sciences. Elsevier, Oxford, pp 13727–13731. Online at http://dx.doi.org/10.1016/B0-08-043076-7/03187-9
Tallis R (2011) Aping mankind: Neuromania, Darwinitis and the misrepresentation of humanity. Acumen Publishing, Durham
Tansey EM (1990) George Eliot's support for physiology: the George Henry Lewes trust 1879–1939. Note Rec Roy Soc Lond 44(2):221–240
Tansey EM (1992) "… the science least adequately studied in England": physiology and the George Henry Lewes studentship, 1879–1939. J Hist Med All Sci 47(2):163–185
Temkin O (1963) Basic science, medicine, and the romantic era. Bull Hist Med 37:97–129
Thompson E, Varela FJ (2001) Radical embodiment: neural dynamics and consciousness. Trends Cogn Sci 5:418–425
Tjoa HG (1977) George Henry Lewes: a Victorian mind. Harvard University Press, Cambridge, MA
Tyndall J (1874) Address delivered before the British Association assembled at Belfast: with additions. Longmans, Green and Company, London
Uffink J (2001) Bluff your way in the second law of thermodynamics. Stud Hist Phil Mod Phys 32:305–394
Vignoli T (1885) Myth and science: an essay, 3rd edn. Kegan Paul, London
Wilson M (2002) Six views of embodied cognition. Psychon Bull Rev 9:625–636
Wozniak RH (1992) Mind and body: René Descartes to William James. National Library of Medicine/American Psychological Association, Washington, DC
Yolton JW (1984) Thinking matter: materialism in eighteenth-century Britain. Blackwell, Oxford
Zeh HD (2007) The physical basis of the direction of time, 5th edn. Springer, Berlin

Chapter 8
Herbert Spencer: Brain, Mind and the Hard Problem

C.U.M. Smith

Herbert Spencer: letter to his father whilst working on his *Psychology* (Duncan 1911, p. 75).

Nineteenth-century thought was pervaded by the idea of progress, by a new consciousness of historical development, by theories of evolution. The task of integrating these ideas formed the life-work of the British philosopher, Herbert Spencer (1820–1903). In the closing decades of the nineteenth century he was regarded as a major thinker.[1] Even Charles Darwin, who on other occasions was more than somewhat disparaging, wrote to Ray Lankester that he suspected that 'hereafter he [Spencer] will be looked at as by far the greatest living philosopher in England; perhaps equal to any that have lived.'[2] Spencer's self-imposed task was to collate all the disparate thought of the nineteenth century, from cosmology to sociology, into one great synthesis. This synthesis was published as *A System of Synthetic Philosophy* in successive volumes, beginning with *First Principles* in

Author was deceased at the time of publication.
'… in my private opinion it [*The Principles of Psychology*] will ultimately stand beside Newton's *Principia.*' Much of the material of this chapter has been derived from my earlier publications on Spencer's psychology: (Smith 1982a, 1983a, 2007).

[1]According to the *Dictionary of National Biography* Spencer's 'influence in the latter part of the nineteenth century was immense … [his] influence extended throughout the world.' The popularity of his publications may be judged by the fact that he was offered some 22 academic distinctions, ranging from university doctorates to fellowships and presidencies of learned societies all over the world. These distinctions, moreover, were offered in spite of it being known that he was not interested in honours such as these and, indeed, habitually declined them.

[2]Letter, March 15, 1870, in Darwin 1887, vol. 3, p. 120.

C.U.M. Smith (✉)
Emeritus, Department of Vision Sciences, Aston University, Aston Triangle,
B4 7ET Birmingham, UK
e-mail: a.m.jennysmith@gmail.com

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_8,

1862 and ending with *Principles of Sociology* in 1896. On dictating the final words of the final volume of his great synthesis, he remarked to his amanuensis, ‘It is for this I have lived’.

Accounts of Spencer’s early life may be gleaned from his autobiography,[3] letters[4] and biographies.[5] He was born at Derby in 1820 and, like several other original British nineteenth-century thinkers, received little formal education. In 1837 he gained employment as a civil engineer working on the London to Birmingham railway. This phase of his life did not, however, last long. In 1841, as he wrote in his diary, ‘got the sack – very glad’. For the remaining 62 years of his life (except for a short return to engineering in 1844–1846) he supported himself by journalism and writing. Nevertheless this early experience with the railways may have been seminal. He is said to have had his initial interest in evolution stirred by the geological strata and organic fossils revealed by railway cuttings.

His first publications had to do with sociology and dissenting radicalism. Although, looking back as an 80-year old, he insisted that it was the idea of evolution working itself out through all areas of human knowledge that the *Synthetic Philosophy* was intended to display,[6] we might suspect that his original impulse was to provide some intellectual foundation for radical programs of social change. Old men forget!

However this may be, he soon turned his attention to biology and psychology. In both these subjects he was self-taught. His knowledge of biology was largely derived from reading Carpenter,[7] Milne-Edwards[8] and, through them, von Baer.[9] His guiding idea was derived from von Baer’s embryology. Von Baer had shown that embryological development proceeded from the general to the special, that the chick (to use von Baer’s own example) begins as a vertebrate, differentiates into a gallinaceous bird and, finally, into a domestic fowl.[10] This idea was to form the basis of a famous illustration in Haeckel’ *Anthropogenie* where he showed that a number of vertebrate embryos are hardly distinguishable to begin with but, as development proceeds, differentiate into fish, salamander, turtle, chick, pig, cow, rabbit and human.[11] Looking back, in mid-career, Spencer remarked that ‘that which really has exercised a profound influence over my thought is the truth … put into definite shape by von Baer – the truth that all organic development is a change from a state of homogeneity to a state of heterogeneity … [this] formula of von Baer acted as an organizing principle’.[12]

---

[3] Spencer 1904.

[4] Duncan 1911.

[5] Webb 1979, pp. 21–38; Peel 1971; Francis 2007.

[6] Letter…, Duncan 1911, p. 546.

[7] Carpenter 1839.

[8] Milne-Edwards 1834.

[9] von Baer 1828, 1837.

[10] Von Baer 1828, p. 140.

[11] Haeckel 1874.

[12] Spencer 1864.

It is easy to see that von Baer's embryological concepts form the kernel of Spencer's famous definition of evolution:

> Evolution is an integration of matter and a concomitant dissipation of motion; during which matter passes from a relatively indefinite, incoherent homogeneity to a relatively definite coherent heterogeneity; and during which the contained motion undergoes a parallel transformation.[13]

Spencer would have found this generalization fully applicable to the comparative anatomy discussed in Milne-Edwards's and Carpenter's zoology. Progress from the simplest forms of multicellular animal life, for instance the Cnidaria, to the more advanced forms found amongst the arthropoda and cephalopod mollusca, seemed to show just this 'progress' from homogeneity to heterogeneity. Carpenter, indeed, makes explicit reference to Milne-Edwards and von Baer, writing that 'whether we trace the 'Archetype' of each great subdivision of the animal kingdom… or whether we follow any organ or system, we recognize one and the same plan of progression, namely, *from the general to the special*… this idea, due to von Baer, is developed in a very admirable manner by Milne-Edwards…'[14] It is clear that the evolutionary theory on which Spencer based his thought is markedly pre-Darwinian. It stems from the work of early nineteenth-century embryologists. In this way it is quite different from the mechanistic paradigm which pervades the thought of the great seventeenth-century philosophers and their associationist successors.

8.1 Origins of *The Principles of Psychology*

The first major work which Spencer attempted, after vigorously arguing the case for evolution in an article in the *Leader* on the 'Development Hypothesis',[15] was his *Principles of Psychology*. The first edition was published in 1855 and to his chagrin fell almost still-born from the press. In later years he believed that this was largely because the book was years ahead of its time.[16] There is no doubt, however, that the mental effort involved in putting together the book broke Spencer's health. He was never again able to concentrate for long periods on any intellectual task: either reading or writing.[17] All his subsequent works, including subsequent editions of the *Psychology,* were dictated to amanuenses and he was forced to develop stratagems such as rowing on the Regents Park lake and vigorous games of racquets to break up his periods of reading and dictation.

---

[13] Spencer 1870a, p. 396.

[14] Carpenter 1839, p. 20.

[15] Spencer 1852.

[16] Duncan 1911, p. 140.

[17] In an interesting article Martin Raitier has suggested that the effort of writing the *Psychology* brought on the reading epilepsy from which he suffered for the rest of his life (Raitiere 2011).

The sources of the *Psychology* are difficult to discern. Spencer gives few references in the text and is known to have been a somewhat irascible reader.[18] In his *Autobiography* he gives the impression that it had sprung forth fully armed from his brain.[19] This can hardly have been the case. Spencer had been occupied with psychological speculation from at least the early 1840s, 10 years before the first edition of the *Psychology* was published. In his twenties he had been much interested in phrenology and had written several papers which were published in the *Zoist*.[20] In the early 1850s, immediately before writing the *Psychology*, he had been intimate with George Henry Lewes who, at that time, was presenting lectures on physiological psychology at Finsbury. In his *Autobiography*, written nearly 50 years later, he recalls many animated discussions with his friend. But, according to Spencer's own account, the seed which caused his ideas to crystallize was John Stuart Mill's *System of Logic*. This triggered a lengthy article entitled 'The Universal Postulate' which was published in the *Westminster Review* of 1853.[21]

8.2 The 'Universal Postulate'

The 'Universal Postulate' forms the first part of the first edition of the *Psychology* but is relegated to the second volume in the second and subsequent editions. This is due to Spencer's perception that the 'analytic approach', of which the postulate forms the basis, is 'much less readable than the synthetical.'[22] It is, consequently, the latter approach which forms the first volume of editions after the first.

The 'Universal Postulate' formed Spencer's 'Archimedean point'. Like many philosophical thinkers (one thinks of Descartes and his process of hyperbolic doubt), Spencer was not prepared to begin without establishing a firm foundation – throwing away loose earth and sand, to quote Descartes again, to reach solid rock. 'No rational Psychology', he writes in the first edition of the *Psychology*,

> can be constructed save on the basis of some acknowledged relation between thought and the subject matter of thought – between mind and nature. No explanation whatever can be given to any act of intelligence, but what implicitly affirms or denies certain ontological propositions. Hence, unless some such proposition can be established, no superstructure of science is possible.[23]

Spencer's definition of this foundational concept – the 'Universal Postulate' – emerges from a critical review of the epistemologies of Reid, Mill, Berkeley, Hume,

---

[18] He was fond of relating how he had thrown down a copy of Kant in disgust after finding that he disagreed with the first two or three pages (Elliot 1975).

[19] Spencer 1904, vol. 1, p. 391: 'the data for the subjective part, which was after a manner unlike that commonly adopted , were lying ready internally: the views taken in the objective part were so alien to those of preceding psychologists that no extensive study of their writings was necessary'.

[20] Spencer 1844a, b, c.

[21] Spencer 1853.

[22] Spencer 1855, Preface, p. iv.

[23] *Ibid.*, p. 34.

Kant and Mansel. It is only after finding them all, for one reason or another, defective that he proposes his own bedrock principle: the notion of 'belief'. He supports this conclusion by arguing that

> to say… there is no belief, is to utter a belief which denies itself – is to draw a distinction between that which is, and that which is not, and at one and the same time to say, do not distinguish between that which is and that which is not.[24]

What did Spencer mean by *belief*? In his 1853 article in the *Westminster Review* he defines it in the following way: 'Every logical act of the intellect is a predication – is an assertion that something is; and this is what we call belief'.[25] Unlike earlier epistemologies which ultimately founded themselves on observation – 'there is thought now, therefore … (*si enim fallor, sum; cogito ergo sum*), or Locke's method of 'looking into his own mind and seeing how it wrought'[26] – Spencer, like Goethe's Faust, founded his system on an *act*. It may be that this shift from spectatorship to participation is associated with the acceptance of a fully evolutionary philosophy.[27] This developing vision of man as part of nature is, of course, fully shared by Spencer's great contemporary, Charles Darwin. In the *C Notebook*, for instance, which he kept immediately after his return from the *Beagle* circumnavigation, Darwin writes, 'Man in his arrogance considers himself a great work worthy of the interposition of a Deity. More humble and I believe truer to consider him created from the animals.'[28]

But, returning to Spencer, we have, of course, all sorts of beliefs, some more reliable than others. It is needful, as Spencer would say, to search out and classify the belief, or beliefs, of which we can be most certain. By reviewing the huge variety of beliefs which can be entertained in the mind, Spencer believed he could discern a class which it is impossible to doubt. We have no choice in the matter: their negation is inconceivable.

This defines his 'Universal Postulate'. 'Knowledge of the highest validity' is that of which the negation is inconceivable. Spencer gives many examples of this grade of knowledge. 'It is inconceivable' he writes,

> that one side of a triangle is equal to the sum of the other two sides; the two sides cannot be represented in consciousness as being equal in joint length to third side, without the representation of a triangle being destroyed; and the concept of a triangle cannot be framed without a simultaneous destruction of a concept in which both these magnitudes are represented as equal. That is to say, the subject and the predicate cannot be united in the same intuition – the proposition is unthinkable.[29]

[24] Spencer 1853, p. 519.

[25] *Ibid.*, p. 518.

[26] Locke 1690, Book 1, chapter 1.

[27] This shift from the passive to the active, from spectatorship to involvement, can also be found in the evolutionary neuropsychology published by Erasmus Darwin at the end of the eighteenth and beginning of the nineteenth centuries (see Smith 2005).

[28] Darwin 1838, pp. 196–7.

[29] Spencer 1865, p. 535.

In this example the Postulate is being used to define analytic or, as Whewell termed it, necessary truth.[30] But Spencer maintains that the Postulate shows that there are other sorts of belief that hold equally exalted places in the hierarchy. 'Whilst looking at the sun' he writes, 'a man can no more conceive that he is looking into darkness than he can conceive that the part is greater than the whole'.[31] He gives several examples to show that what Herbert Feigl calls 'raw feels'[32] and Russell 'egocentric particulars' are also to be classified as 'knowledge of the highest validity'. These propositions remind us, of course, of the *cogito*. It is impossible to believe, when warm, that this sensation is not occurring.

The use of the Universal Postulate leads Spencer to adopt a position which he terms 'transfigured realism'. He argues, to his own satisfaction, that all other positions are less certain. In essence he agrees with Wittgenstein that doubt is parasitic upon certainty: 'If I want the door to turn, the hinge must stay put'.[33] Of Hume he writes that 'to conclude that there is no proof of an external world is to reason my way to the conclusion that reason is fallacious.'[34] Of Berkeley's *esse est percipi*, that 'all those bodies which compose the mighty frame of the world have not any substance without a mind, that their being is to be perceived or known..,' he similarly comments, 'How can we be sure of this?' Does not Berkeley's argument ultimately 'base upon a thing's existence the proof of its non-existence'? Such arguments, he concludes, are 'like many kindred kinds, self-destructive; [they] repeatedly assume the validity of that whose validity [they] question.'[35]

In contrast, use of the 'Universal Postulate' leads, Spencer argues, to a position which, as we noted above, he dubs 'transfigured realism'; than which, he continues, no position is more certain. In all other foundational positions:

> …the derived is to set aside that from which it is derived; a series of links is to be regarded as stronger than any one of its single links; and consciousness is more trusted when its terms are indistinct than when they are distinct.[36]

Spencer's notion of transfigured realism is explained most fully in the second volume of the second edition of *The Principles of Psychology*.[37] In essence he distinguishes it

[30] *Ibid*., p. 521n: '…Dr Whewell defines necessary truths as "those in which we not only learn that the proposition *is* true, but see that it *must* be true; in which the negation of the truth is not only false but impossible…'.

[31] Spencer 1855, p. 28.

[32] Feigl 1958, p. 28: 'Don't you want anesthesia if the surgeon is to operate on you? And if so what [you] want prevented [is not behavior but] the occurrence of the very raw feels of pain, is it not?'

[33] Wittgenstein 1975, §343.

[34] Spencer 1855, p. 42.

[35] *Ibid*., p. 24.

[36] Spencer 1872, vol. 2, p. 490.

[37] *Ibid*., §472–473; see also Smith 1983a, b.

from 'crude realism' by asserting that 'reality' is ultimately unknowable and that our perceptions only provide us with understandings which evolution has proved to be beneficial to creatures such as ourselves.[38]

8.3 The Experience Hypothesis

The Universal Postulate is, Spencer maintains, itself knowledge of the very highest validity: 'not even a reason for doubting its validity can be given without tacitly asserting its validity.'[39] It is important to note the terminology. Spencer speaks of knowledge of the 'highest validity', not *a priori* truths. His philosophy, he insists, is founded on what he called the 'experience hypothesis'. When he published the first edition of the *Psychology* in 1855, he believed that he had found a way of reconciling this hypothesis with 'its antagonist hypothesis of forms of thought'.[40] This reconciliation, he writes, is effected by an application of evolutionary theory to the ancient controversy. Such an application, he continues, 'furnishes a solution to the controversy between the disciples of Locke and those of Kant'.[41]

In other words, Spencer, in the middle of the nineteenth century had hit on the evolutionary explanation of the Kantian categories subsequently popularised in the twentieth century by Popper, Campbell, Lorenz and others. It is not, says Spencer, the experience of the individual that provides the peculiar certainty of the Kantian categories or of the axioms of geometry, but the cumulative and inherited experience of the countless lives that have gone before. In the 1872 second edition of the *Psychology* he expresses this view with great force: 'Space-relations have been the same not only for all ancestral man, all ancestral primates, all ancestral orders of mammals, but for all the simpler orders of creatures. These constant space-relations are expressed by definite nervous structures, congenitally framed to act in definite ways and incapable of acting in different ways. Hence the inconceivableness of the negation of a mathematical axiom, resulting as it does from the impossibility of inverting the actions of correlative nervous structures, really only stands for the infinity of experiences that have developed these structures.'[42]

At bottom, then, analytic truths are, like the synthetic truths of empiricism, derived from experience: the former derived from the accumulated experience of an infinity of ancestors, the latter from the experience of the individual, especially of the infant during early development.

---

38 For a modern version of this position see Metzinger 2009.

39 Spencer 1872., vol. 2, p. 491.

40 Spencer 1855, p. 23n; Spencer 1864, vol. 2, p. 413.

41 Spencer 1855, p. 578.

42 Spencer 1872, p. 419.

8.4 Subject and Object

Having 'cast aside loose earth and sand' and found a sound base, Spencer now felt able to set about developing a psychobiology in which the categories of 'subjectivity' and 'objectivity', of 'self' and 'not self', emerge from the initially undifferentiated 'chaos' of the phenomenological field. Instead of beginning his meditation by sequestering himself in a Bavarian poêle, Spencer chose a much more English scene: sitting on a bench on a seaside promenade with the sea breeze blowing in his face, the breakers crashing on the pebbly shore, aware of the spray-filled air and the smell of seaweed, with the overcast occasionally parting to reveal a white-cliffed headland and groups of anchored boats.[43]

'Bracketing out' all preconceived knowledge and focusing on the phenomena primarily 'given' reveals, he says, two sets of events which may be distinguished most readily by their comparative 'vividness'. The more vivid events – the blue, the white, the crash and rumble, the coolness and damp, the odour, the pressure (to assign these 'raw feels' their received names) – differ from the less vivid occurrences – memories, associations, anticipations (to give them, once again, their received names) – in a number of ways.[44] Spencer provides a list. The first sequence of happenings, the vivid sequence, seem to be quite independent of the second sequence, the faint occurrences. The opposite is not the case. The vivid events, 'sweeping past', as he puts it, seem to drag the faint states along with them, as if by a process of lateral' adhesion'.[45] 'The sounds of the pebbles rolled about by the waves,' he remarks by way of example, 'inevitably draw along ideas of shapes and colours and hardnesses'. Second, while the antecedent to any event in the first sequence may or may not be present in the stream of happenings – the sudden appearance of a dog's bark has, for instance, no discoverable antecedent in the sequence of vivid events – the antecedent to an event in the second sequence is always, in principle, to be found in one of the earlier events in that sequence. Spencer goes on to distinguish a number of other differences between these two trains of happenings. He concludes by arguing, rather as Piaget was later to argue,[46] that during the first few months and years of our lives, the consistent difference between these two sets of events leads to one of the most deeply embedded of our categorisations – that between the 'objective' and the 'subjective'.

In all this Spencer is very close to much subsequent phenomenological thought. When our 'received' understanding is stripped away, we are left with the 'phenomenal field', the 'this-here-now'. But so far, Spencer has taken the stance of an uninvolved spectator. He has described that which is 'disclosed' as if the sights and sounds and odours, the memories and anticipations, presented themselves, willy-nilly, as if, to use Whitehead's phrase, they were an uncontrollable 'stream of happenings'. But

[43] *Ibid*., pp. 454–455.

[44] *Ibid*., pp. 463–4.

[45] *Ibid*., p. 459.

[46] Piaget 1972.

this, we all know, is not the case. We are not merely observers, but inextricably involved. We are aware of a power which 'wells up' and which we can control. In essence, according to Spencer, this is the sense of muscular effort. It leads to one of the basic axioms of Spencer's neurophilosophy: the notion of '*force*': 'the ultimate of ultimates.'[47] The sense of touch and the resistance touch seemed to Spencer to be one of the most basic of all the phenomena which present themselves in the stream of happenings. Tactile sensations, resistances, are everywhere; whilst we live we are immersed in them whether we stand, sit or lie down. We saw in chapter 1 that Aristotle is at one with Spencer in taking touch to be the primary sense. Spencer writes in a similar vein 'Excluding the lowest animals… there are none (Aristotle would have disagreed here) but have, at every moment of their lives, some impression of resistance'. Such impressions, he continues, 'form, as it were, the weft of that tissue of thought which we are ever weaving'.[48]

Throughout the *Psychology* Spencer returns again and again to the ontological question. He is quite clear that, in the final analysis, our experience is precisely *our* experience. The notions of 'subjectivity' and 'objectivity' develop as a fundamental classification of our experience of 'being in the world'. 'The normal processes of thought,' he writes, 'inevitably originate this inexpressible but indestructible consciousness of existence beyond the limits of consciousness; which is perpetually symbolized by something which is within its limits.'[49] The world, he insists, is ultimately a mystery; our situation, as existentialist writers frequently remind us, ultimately unfathomable.[50]

Spencer was sensitive to this unfathomability. 'Consciousness' for him, as for Sartre, is something one can only catch sight of out of the corner of the eye. 'The consciousness which says "I think"' writes Sartre. 'is precisely not the consciousness which thinks'.[51] The latter is always ahead of thought, which looks back an instant later. It is, to use another Sartrean expression, a 'clear wind'. Spencer writes as follows:

> Be the thing contemplated in the act of cognition a symbolized activity existing beyond the Mind, or be it a past state of the Mind itself, that which contemplates it is distinct from it. Hence were it possible for the substance of Mind to be present in any state of Mind, there would still have to be answered the question – What is it then that contemplates it and knows it?… The substance of Mind escapes into some new form in recognizing some form under which it has just existed… In brief, a thing cannot at the same time be both subject and object of thought; and yet the substance of the Mind must be this before it can be known… Mind remains unclassable and unknowable.[52]

---

[47] Spencer 1870a, p. 169; and in the second edition of the *Psychology* (vol. 2, p. 232) he writes '… the impression of resistance. This is the primordial, the universal, the ever-present constituent of consciousness'. The idea that the physical concepts of force and energy are derived from our conscious experience of muscular energy is familiar today from the work of historians of science such as Max Jammer (Jammer 1957, p. 7).

[48] Spencer 1872, p. 233.

[49] *Ibid*., p. 488.

[50] *Ibid*., 503.

[51] Sartre 1957, p. 45.

[52] Spencer 1870a, b, p. 147–8.

8.5 Spencer and Descartes

Spencer's tactile neuropsychology takes him far from the optical version we have learnt from Descartes and, further back still, from Plato and Euclid.[53] The contrast is instructive. The fundamental feature of the 'external world' for Descartes was 'extension'. In the *Principles of Philosophy* he writes:

> If, wherever our hands moved in a given direction, all the bodies in that area were always to retreat with the same speed as our hands approached, we should never have any sensation of hardness. Now it is inconceivable that, if bodies did retreat in this way, they would thereby lose their nature as bodies; so this nature cannot consist in hardness … The nature of matter, or of body considered in general [consists] … simply in its being a thing that has extension in length, breadth and depth.[54]

Spencer has it quite the other way around. For him the notion of 'extension' is derived from the experience of 'resistance'. Spencer, compared with Descartes, is immersed in the world of events, feeling them on his pulse; Descartes, in contrast, in pre-evolutionary times, seems to be a spectator, watching bodies retreat as he stretches out for them, uninvolved, standing over against the world.

This contrast between Cartesian spectatorship and Spencerian involvement becomes evident in yet another central part of Descartes' metaphysics. As is well-known, Descartes had to appeal to a version of St Anselm's ontological proof to assure himself of the truthfulness of his perceptions. For without the proof that a just and beneficial God existed, the reports of the senses might be mere illusions. St Anselm's proof depends crucially on the proposition that, to quote Descartes. '… it is no less contradictory that the perfect should follow from and depend on the less perfect, than that something should proceed from nothing.'[55] In other words it is impossible for a being lower in the *scala naturae* to generate one higher in that scale. Furthermore, the idea of a perfect being cannot be conceived by an imperfect being. The idea of God must, therefore, have been implanted in the human mind. As Alexander Koyré wittily writes, corresponding to the *cogito*, 'I think therefore I am', there is another Cartesian axiom, 'God is thought of, therefore God exists'.[56]

These propositions are fragments of a world-view totally alien to the world-view of Spencer's nineteenth-century evolutionism. They presuppose a 'great chain of being', to be sure, but it is a static chain (see chapter 11) and influences run from the top to the bottom, from the great ones in Heaven (or Earth) to lesser mortals in society or the animal kingdom. Spencer's evolutionary philosophy envisages a movement in quite the opposite direction, from fish to philosopher. The evolutionist inhabits a different world, a world in which it is not necessary to continually attempt

[53] Francis Bacon observes that 'God hath framed the mind of man as a mirror or glass, capable of the image of the universal world' (Bacon, 1605, *Advancement of Learning*, Book 1, §3).

[54] Descartes 1664, Part 2, §4.

[55] Descartes 1637, chapter IV, p. 33.

[56] Koyré 1970, p. xi.

to find proofs against solipsism, a 'scandal' according to Martin Heidegger[57]: humans are, from the first, part of the evolutionary process. The evolutionary philosophy leads directly to one of the puzzling aspects of the 'hard problem': how and when does consciousness commence, and how is it related to the material within our skulls?

8.6 Neurophysiology and the 'Hard Problem'

The second edition of the *Principles of Psychology* starts with an account of the nervous system – its structure, functions and evolution. The transition to an introspective psychology is made by way of a discussion of 'aestho-physiology', or, in our terms, physiological psychology. 'Feeling and nervous action', Spencer asserts, 'are the inner and outer faces of the same change'.[58] Indeed, in a passage reminiscent of Huxley's phrase 'the mechanical equivalent of consciousness',[59] Spencer writes 'Is there such a connection between a physical change in the nervous system and the psychical change accompanying it, that we may regard the one as the equivalent of the other, the same sense as we may regard so much heat as the equivalent of so much motion?'[60] And Spencer, in this so much more circumspect than Huxley, concludes after some further analysis that there does indeed exist some such quantitative relationship between nervous action and subjectively experienced feeling.

What is this relationship? Spencer gives a fresh twist to the physiological psychology developed by his associationist predecessors. True to his understanding of the basic nature of consciousness, he attempts to dissect out 'unit' feelings and to relate these to 'unit' events in the brain. He argues, accordingly, that

> there is at least one kind of feeling which, as ordinarily expressed, seems elementary that is definitely not elementary. And after resolving it into its proximate components, we can hardly help suspecting that other apparently elementary feelings are also compound, and may have proximate components like those which we can in this one instance identify.[61]

Spencer is referring to his theory of musical sound. He points out that, below a frequency of about 16Hz, sound is experienced as series of individual 'taps' or percussions. Above this frequency our sensation changes and we experience a

[57] Heidegger 1962, p. 249: 'The "scandal of philosophy" is not that this proof [that for an external world] has yet to be given but that *such proofs are expected and attempted again and again…* such expectations, aims and demands arise from an ontologically inadequate way of starting with *something* of such a character that independently *of it* and "outside" *of it* a "world" is to be proved present-at-hand.... If Dasein is to be understood correctly, it defies such proofs, because, in its Being, it already *is* what subsequent proofs deem it necessary to demonstrate for it.'

[58] Spencer 1870a, b, p. 128.

[59] Huxley 1870, vol. 1, p. 191.

[60] Spencer 1870a, b, pp. 116–117.

[61] *Ibid*., 1870, pp. 148–9.

continuous tone. The quality of the tone changes as the frequency changes. Furthermore, the timbre of the tone, its harshness or sweetness, its liquidity or clearness, depends on combinations of frequencies. This shows, Spencer concludes, that

> an enormous number of qualitatively contrasted kinds of consciousness that seem severally elementary, prove to be composed of one simple kind of consciousness, combined and recombined with itself in multitudinous ways.[62]

It is unfortunate for Spencer that by 1870 Helmholtz had already shown that the cochlea did not (and we now understand could not) signal high-frequency tones to the brain in the form of matched impulse frequencies in auditory nerve fibres. Helmholtz's 'place theory' of frequency detection, first proposed in 1863, undercuts the neurophysiology of this part of Spencer's theory. For Spencer wants to argue that just as our perception of music and the vast ranges of conscious states associated with music are built from innumerable combinations and permutations of elementary 'unit shocks', so it is with all our other sensations and emotions.

Indeed, he wants to go further. He wants to argue that there may well be an elementary unit of consciousness, a universal 'atom', from which all the nearly infinite variety of mental states are built. And, of course, he believes he knows what this 'atom' is. It is a sudden 'happening', a sudden 'shock'. He maintains that a sudden nearly instantaneous alteration in sensory input, whether through touch, eye, nose or ear – provided it does not last long enough 'to admit of its being contemplated' – is only experienced as an 'event':

> It cannot be classed as of this or that kind; and becomes a momentary modification very similar to momentary modifications otherwise caused. It is possible, then – may we not say even probable – that something of the same order as that which we call a nervous shock is the ultimate unit of consciousness.[63]

True to his aestho-physiological convictions, Spencer sees the other side, the physical side, of this 'unit of consciousness', this sudden 'shock', as the nerve impulse or, as he calls it, the 'intermittent wave of the nerve current'. The various states of consciousness, the multitudinous different qualities and modes of feeling, are to be correlated with the complex interactions of nerve currents in the indifferent material of the brain. This explains, he writes, the otherwise baffling observation that nerve centres acting as 'seats of different feelings' are anatomically indistinguishable.[64] Once again we are aware of Spencer anticipating puzzlements of a much later era. Braitenberg in his well-known work of cerebral microanatomy emphasized how the histology of the neocortex is 'invariant irrespective of local functional specialisation' and Sur and colleagues have shown that ferret cortex can be 'rewired' so that dedicated auditory areas can serve a visual function.[65]

Spencer thus believes that it is possible, at least in principle, to show that all states of consciousness, however infinite their variety, are but the 'subjective' face of complex

[62] *Ibid*., 1870, p. 150.

[63] *Ibid*., p. 151.

[64] *Ibid*., p. 154.

[65] Roe et al. 1992.

permutations and combinations of identical elementary units whose 'objective' face is the 'intermittent wave of the nerve current'. No single neuron, he argues, corresponds to an external object, anticipating by a century the modern controversy over the existence of 'grandmother' or 'cognitive' cells. Rather, he says, an object is represented by excitement of a whole assembly of neurons, and activity in a different assembly represents each discriminable object and each position and orientation of that object.[66] In this respect, he goes on, the brain may likened to a piano. If the keys of a piano are struck separately about a hundred different notes are elicited; but, consider, he says, the number of different chords which can be obtained when more than one key is struck at once. This number, he calculates, exceeds 10^{30} when all combinations of fifty notes are sounded. Similarly, he says, for the brain: 'a limited number of fibres and cells become the seat of a relatively unlimited number of perceptions.'[67]

Spencer is, however, perfectly clear that it would be absurd to suppose that consciousness is the 'subjective aspect' of each and every nerve current that runs through the brain.[68] In order to proceed further with Spencer's psychobiology it is necessary to examine his answer to Darwin's question: 'How does consciousness commence?'

8.7 Evolution and the Hard Problem

As a convinced evolutionist Spencer has no hesitation in arguing that 'If the doctrine of Evolution is true, the inevitable implication is that Mind can be understood only by observing how Mind is evolved.'[69] In both the first and second edition of the Psychology, he shows how the objective features of mentality grade insensibly from protozoan to philosopher. How do we know this? Spencer answers that we know this with as much and as little certainty as we can attach to our knowledge of the existence of other human minds. If we once grant that other human minds exist, we cannot, he says, deny that animal minds exist.[70]

Spencer's general position is summed up in the *Autobiography* he worked on at the end of his life and which was published posthumously in 1904: 'the form of life which we call Mind, emerges out of bodily life.'[71] He argues that the phylogenetic series shows that animals are continuously adjusting their 'internal relations' to the 'external relations' of the environment. Nowadays we might reword this proposition and say that the animals continuously refine an internal *model* of the circumambient world. Spencer goes on to argue that as animals ascend the *scala naturae,* the internal model becomes more and more accurate, more and more complete, extending from the immediate spatial and temporal environment to embrace wider and wider perspectives.

---

[66] Spencer 1870a, b, p. 562.

[67] *Ibid*., p. 563.

[68] *Ibid*., p. 104.

[69] *Ibid*., p. 291.

[70] *Ibid*., p. 98.

[71] Spencer 1904, vol. 1, p. 470.

Moreover, Spencer argues, the evolution of the nervous system provides yet another instance, and an important one, of his general evolutionary law: the progress from a state of indefinite, incoherent homogeneity to a state of definite, coherent heterogeneity. Nerve centres are at first poorly organized and information processing (to use modern terminology) is difficult and time-consuming. Later, with repetitive use, the neuronal structure of the centres becomes more highly differentiated and thus more effectively organized. The cells of the differentiated structure become more tightly interconnected and the whole centre acts in a more coherent manner. Spencer gives a number of examples of this 'evolution'. They mostly refer to ontogenesis. This cannot surprise us for, as we saw in the first section of this chapter, Spencer's concept of evolution derived from von Baer's embryology, not from Charles Darwin's natural history. Thus Spencer instances the development of speech in human infants. Speech, he says, is learned slowly and with difficulty. This implies, he continues, that the nerve centres responsible for articulatory behaviour are poorly differentiated and poorly interconnected: 'The concomitant sentient states [are therefore] vivid, and, for the moment, all-embracing.'[72] Later, however, speech becomes easy, well-formed and automatic. The responsible nerve centres are, by this token, highly differentiated and their units coherently organized. 'Nerve energy' can consequently flow through these centres with little resistance and is directed outwards without difficulty to the appropriate muscles. Consciousness, the other side of the coin, loses its early intensity and lapses to the low levels associated with automatic activity.[73] Some have subsequently accused Spencer (with some justification) of proposing a 'frictional' theory of consciousness.[74]

Spencer, however, does not necessarily believe that the 'seat of consciousness' is located in the nerve centres which govern behaviour such as articulation. Rather, he believes that consciousness is the accompaniment of activity in 'higher centres'. He argues that early on in the developing nervous system a poorly organized nerve centre is unable to cope with the rush of nerve currents flooding into it. It does not, in this early phase of its development, have the well-structured interconnections between differentiated elements which it acquires later in life. In consequence some of this 'nerve energy', instead of flowing out towards, say, the articulatory muscles, escapes 'centripetally' to a 'higher centre' thus 'awakening a feeling.'[75] Connected with this conclusion is the observation that feelings always have a temporal dimension. He remarks, following Huxley and others, that 'a flash of lightning' is instantaneous, yet the visual sensation lingers on. This, according to his theory, is

[72] Sp, Spencer 1870a, b, p. 106.

[73] *Ibid.*, p. 560.

[74] It is, of course, far from obvious why mere difficulty of discharge, 'friction', should be a necessary and sufficient concomitant of consciousness. Romanes was quick to see this: 'I think, however, that Mr Spencer is not sufficiently explicit … in showing that "the raw material of consciousness" is not necessarily constituted by mere *complexity* of ganglionic action. Indeed, as I have said, such complexity in itself does not appear to have anything to do with the rise of consciousness, except in so far as it may be conducive to what we may called ganglionic friction which may be expressed by delay in response' (Romanes 1883, p. 74n).

[75] *Ibid.*, p. 106.

due to the massive sensory inflow saturating the 'lower centre' so that an overflow of nervous energy surges up a centripetal pathway to a 'higher centre' and, reverberating there, forms the physical aspect of the experience.[76]

But what did Spencer mean by the term 'higher centre'? His concept remains quite acceptable today. 'Higher centres are', he writes, 'those parts of the brain to which information from several or all sensory modalities is sent, co-ordinated and correlated.'[77] At first reading, Spencer seems to be caught in some difficulty here. On the one hand he wishes to assert that these centres represent the climax of the evolutionary process. On the other, his 'frictional theory' of the physical correlatives of consciousness has it that they must be poorly differentiated, incoherent, their elements feebly interconnected. This, far from fitting his definition of an advanced stage in evolution, describes the exact opposite: a primitive beginning. How can he reconcile these two seemingly contradictory positions?

The answer to this apparent contradiction is to be found in the overarching evolutionary theory which forms the setting for Spencer's neurobiology. We have to return, once again, to his grounding in von Baer's biogenetic law and Carpenter's zoology. In *The Principles of General and Comparative Physiology* Carpenter writes:

> In tracing the progressive complication of the psychical manifestations during the early life of the human being, a remarkable correspondence may be observed with the gradual increase in endowments which is to be remarked in ascending the Animal scale.[78]

In short, ontogeny recapitulates phylogeny. The 'higher centres' of lower forms in a phylogenetic series should be seen as homologues of the 'lower centres' of forms higher up the series. In evolutionary 'advance' the 'lower centres' are progressively transformed from an 'incoherent homogeneity' into a 'coherent heterogeneity' as Spencer's theory requires. The 'seat of consciousness' consequently moves from these lower centres upwards to higher centres which have not yet transformed into the smooth automaticity of precisely interconnected heterogeneity.[79] The lower centres are left as 'fully-evolved' ganglia containing populations of well-organised elements where nervous energy normally circulates smoothly so that the 'frictional' conditions which, according to the theory, are the concomitants of consciousness, do not occur. In this way Spencer attempts to provide an answer to evolutionist's abiding questions. Although he does not attempt to explain how consciousness 'commences', he does indicate that, according to his theory, consciousness must extend widely throughout the animal kingdom.

But in spite this belief in the wide extension of consciousness, Spencer contends that in 'lower forms' it is 'vague and unorganised'.[80] It is not clear what reasons he has for this belief other than commonplace disdain for the more primitive members

---

[76] *Ibid.*, p. 107–8.

[77] *Ibid.*, p. 105.

[78] Carpenter, …, p. 458.

[79] Spencer 1870a, b, p. 105.

[80] *Ibid.,* p. 507.

of the animal kingdom.[81] When he turns his attention to some of the lower invertebrates, he wants to say that their consciousness is sectionalized like their anatomy. The 'ganglionic consciousness' of annelid worms and insects seems to him to support, once again, his evolutionary law. A lowly form, he writes, on seeing its customary prey approaching, will have built up in its 'higher centres' pools of neuronal excitement which cannot be released until the appropriate consummatory act of snapping up the victim has been achieved. The anticipatory excitation of the ganglionic elements, before discharge is possible, constitutes, he writes, the physical side of the perception.[82] But this anticipatory activity is largely reactive. The predator brings little foresight to the killing. 'Higher' forms escape this automaticity. Increasingly, as we noted above, they 'model' the environment. Spencer's theory of how these modelling mechanisms might work in the higher animals involve interesting analogies with the 'tune-boards' of the piano-mécanique first demonstrated at the Great Exhibition of 1851.[83] He is groping for technological analogies, which in the middle of the nineteenth century did not yet exist. Nevertheless, this dawning notion of internal programs, subroutines and so on marks a significant advance over Descartes' hydraulically powered reflex automata. It would, however, take us too far from the subject of this chapter to follow his lucubrations further.[84]

8.8 Panpsychism?

'Organisms' Spencer asserts, and a century and half of subsequent research has gone far to confirm the assertion, 'are highly differentiated portions of the matter forming the Earth's crust and its gaseous envelope; and their differentiation from the rest has arisen, like other differentiations, by degrees'.[85] Yet there is a glaring problem. Darwin's friend and co-discoverer of evolution by natural selection, Alfred Russel Wallace, put the problem succinctly:

> If a material element or a combination of a thousand material elements in a molecule, are alike unconscious, it is impossible for us to believe that the mere addition of one, two or a thousand other material elements to form a more complex molecule could in any way tend to produce a self-conscious existence… There is no escape from the dilemma, either all matter is conscious or consciousness is, or pertains to something distinct from matter.[86]

Has consciousness somehow 'appeared', a new phenomenon, not hinted at in the world before higher forms of life evolved? Or is it an infinitesimal phenomenon, an unnoticeable whisper, but, nevertheless, existing in the world from the beginning?

[81] We still await a good reason to conclude that the qualia associated with the reactions of animals lower in the *scala naturae* than the vertebrates are significantly less vivid than our own.

[82] Spencer 1870a, b, p. 561.

[83] *Ibid.,* p. 567.

[84] See *Ibid*., pp. 564–71 and account in Smith 1982a, b.

[85] *Ibid.,* pp. 137–8. For a recent assessment see Rogers 2012.

[86] Wallace 1870, p. 209.

Is it to be compared with gravitational attraction, hardly to be detected until matter has aggregated into sufficiently massy bodies, perhaps, in the case of consciousness, into bodies of a certain minimal degree of integrated complexity?[87]

Spencer struggles to answer these questions. Like Huxley and other scientists of the period, he was anxious to establish that he was not a 'crass materialist' (see chapter 11). 'Were we compelled to choose' he writes (in almost the same words as Huxley used in his similar disavowal), 'between the alternative of translating mental phenomena into physical phenomena, or of translating physical phenomena into mental phenomena, the latter alternative would seem to be the more acceptable of the two.'[88] We saw above that Spencer argued that the mind was built of combinations and permutations of unit 'raw feels'. This was the most fundamental feature of our being. It was from 'living through' these unit qualia that, in the last analysis, we derive our notion of resistance and thence of matter, force and, in a word, of the 'external' world. If, Spencer sums up, the observer 'regards his conceptions of these activities lying beyond the Mind, as constituting knowledge of them, he is deluding himself: he is but representing these ideas in terms of Mind, and can never do otherwise. Eventually he is obliged to admit that his ideas of Matter and Motion, merely symbolic of unknowable realities, are complex states of consciousness built out of units of feeling.'[89]

If, then, we have to admit, with Spencer, that matter and motion are merely terms we give to 'complex states of consciousness', then what are we to make of the 'nerve current' or, in present-day parlance, the 'action potential'? Textbooks illustrate the sharp sigmoid curve on the oscilloscope screen. A deep scientific theory connects this artefact with the opening of ion gates in nerve fibre membranes and the consequent fluxes of ions into and out of the fibre. The whole of this theory is underwritten by a multitude of observations, repeated over and over again. But all, at bottom, rely on noting the qualia 'lived through' by the experimenter. Spencer, making a similar analysis of the fundamentals of neurophysiology as it was understood in the mid-nineteenth century, writes that 'the conception of an oscillating molecule is built out of many units of feeling; and to identify with a nervous shock would be identify a congeries of units with a single unit'[90] And so, on this analysis, is the conception of a 'flux of ions'. It is the shorthand description of the conclusions derived from innumerable experiments; the three-word label of a near-infinity of complex states of feeling. Each observation of a pointer, an oscilloscope trace, a chromatography column or electrophoresis strip is, on Spencer's analysis, a state of consciousness built of a constellation of 'unit feelings'. Thus it would be a very hasty move to identify a 'unit feeling' with a 'nerve current' or 'action potential'. For, as we have noted, the term 'nerve current' itself stands for a 'whole congeries

[87] Smith 1983a, b.

[88] Spencer 1870a, b, p. 159.

[89] *Ibid.*, p. 160.

[90] *Ibid.*, p. 158.

of units of feeling'.[91] It is clear that we are attempting the impossible if we try to 'identify' in any simple and straightforward way a 'unit of feeling' with a 'whole congeries of unit feelings'. Does it make any sense to equate the 'raw feel' of, say, the colour 'red', with the multitudinous 'raw feels' involved in investigating what we call its neurophysiological basis?[92] We have arrived at a *reductio ad absurdum*. Solipsism, it has been remarked, requires not so much a philosopher as a mental ward.

Fortunately, Spencer does not feel forced to choose between what he calls 'idealism' and 'realism'. We have seen that the use of his Universal Postulate leads him to the position of 'transfigured realism' which, he believes, 'accepts from each a moiety… but rejects the rest.'[93] Does this position help us understand 'how consciousness commences'? Do all events have a 'physical' and 'mental' aspect? Wittgenstein called this mere 'image-mongery'.[94] Nonetheless, as we noted in chapter 11, it has attracted many prominent evolutionary thinkers. Spencer also regarded it as a possible hypothesis. 'This hypothesis', he writes, referring to panpsychism, 'remains open'.[95] He does not pursue the matter further. Possibly he felt like Huxley before him that it was an unprofitable speculation when there was so much else to do. He returns, instead, to his consistently expressed belief that we can know no more of the nature of 'unit feelings' than of the nature of 'unit matter': *ignoramus et ignorabimus*. We may analyse both, he argues, into their 'ultimate homogeneous units' and yet remain totally ignorant of what either is. 'Our only course', he concludes, 'is constantly to recognize our symbols as symbols only; and to rest content with that duality of them which our constitution necessitates….. The conditioned form under which Being is presented in the Subject, cannot, any more than the conditioned form represented in the Object, be the Unconditioned Being common to the two'.[96]

8.9 Concluding Remarks

Thomas Huxley, who was a profound and insightful judge of the intellectual abilities of his contemporaries, remarked, in reply to Beatrice Webb's opinion that Spencer had welded his great system from the disjointed theories of the time, that 'he (Spencer) is the most original of thinkers, though he has never invented a new thought'.[97] His ideas seem to have been absorbed by a process of intellectual osmosis from the major debates of the time and to have been set in order by a powerful synthetic intellect.

---

[91] *Ibid.*, p. 158.

[92] For an analogous critique of solipsism see Deutsch 1997, pp. 81–4.

[93] Spencer 1890, Introduction, p. vi.

[94] Wittgenstein 1976, §390.

[95] Spencer 1870a, b, p. 160–161.

[96] *Ibid.*, p. 162.

[97] Webb 1979, p. 28.

His reputation at the end of the nineteenth century was that of a major and original philosophical thinker. However, as we noted at the beginning of this chapter, the dawn of a new century saw that reputation rapidly decline.[98] At the beginning of that century mainstream psychologists turned their attention away from the neuropsychological issues which had so concerned Spencer, to the correlation of animal (and human) responses which could be studied and measured in the laboratory.[99] Watsonian behaviourism swept all before it and so thoroughly reset the psychologist's concept of the nature of their subject that Spencer's work came to be seen as 'unscientific' and obsolete. His concern with the central issue of how the goings-on in the substance of the brain could be related to our everyday lived-through 'raw feels, or qualia, came to be seen as not a proper study for psychologists. It is only in comparatively recent years, with the huge increase in neuroscientific knowledge and the rise of cognitive neuropsychology that the problem has once again come to the fore. Once again the concerns that filled many of the pages of *The Principles of Psychology* have become the subject of intensive debate. As Francis Crick wrote in capital letters towards the end of his last book, the time has come for every laboratory working on the visual system [and, indeed, on other areas of brain science] to have a 'large sign posted on its walls, reading CONSCIOUSNESS NOW'.[100]

References

Braitenberg V (1977) In: Heim R, Palm G (eds) Theoretical approaches to complex systems. Springer, Berlin

Brazier MAB (1959) The historical development of neurophysiology. In: Field J, Magoun HW, Hall VE (eds) Handbook of physiology, section: neurophysiology. American Physiological Society, Washington, DC

Carpenter WB (1839) The principles of comparative physiology. Churchill, London

Crick FHC (1995) The astonishing hypothesis. Simon and Schuster, London

Darwin CR (1838) C Notebook. In: Barrett PH (1974) Metaphysics, materialism and the evolution of mind: early writings of Charles Darwin. University of Chicago Press, Chicago

Darwin F (ed) (1887) The life and letters of Charles Darwin, 3 vols. John Murray, London

Descartes R (1637) Discourse on method (ed and trans: Anscombe E, Geach PT). Nelson, London

Descartes R (1664) Principles of philosophy (ed and trans: Anscombe E, Geach PT). Nelson, London

Deutsch D (1997) The fabric of reality. The Penguin Press, London

Duncan D (ed) (1911) Life and letters of Herbert Spencer. Williams and Norgate, London

[98] Nevertheless, although Spencer's *Psychology* had little overt influence on the psychology in the century after its author's death, its covert influence has been considerable. Translations of Spencer's works into Russian influenced Sechenov and through Sechenov, Pavlov (see Brazier 1959, p. 54) and, in England, exerted a profound effect on John Hughlings Jackson, often regarded as the founder of British neurology (see Smith 1982b). Jackson, in his turn, influenced not only neurology but also Sigmund Freud and Arnold Pick, one of the inaugurators of psycholinguistics in the twentieth century.

[99] See Watson 1913.

[100] Crick 1995, p. 253.

Elliot HSR (1975) Herbert Spencer. Dictionary of national biography. Oxford University Press, Oxford
Fiegl H (1958) The ‘mental’ and the ‘physical’. University of Minnesota Press, Minneapolis
Francis M (2007) Herbert Spencer and the invention of modern life. Cornell University Press, Ithaca
Haeckel E (1874) Anthropogenie oder Entwickelungsgeschichte des Menschen. Engelmann, Leipzig
Heidegger M (1962) Being and time (trans: Macquarrie J, Robinson E). Blackwell, Oxford
Huxley TH (1870) On Descartes discourse touching the method of using one’s reason rightly and of seeking scientific truth. In: Huxley TH (ed) Collected essays. Macmillan, London
Jammer M (1957) Concepts of force. Harvard University Press, Cambridge, MA
Koyré A (1970) In: Anscombe E, Geach PT (eds) Descartes: philosophical writings. Nelson, London
Locke J (1690) An essay concerning human understanding. Printed for A. & J. Churchil, London
Metzinger T (2009) The ego tunnel. Basic Books, New York
Milne-Edwards H (1834) Eléments de zoologie. J.-B. Bailliere et fils, Paris
Peel JDY (1971) Herbert Spencer: the evolution of a sociologist. Heinemann, London
Piaget J (1972) Psychology and epistemology. Penguin, Harmondsworth
Raitiere MN (2011) Did Herbert Spencer have reading epilepsy? J Hist Neurosci 20:357–367
Roe AW, Pallas SL, Young HK, Sur M (1992) Visual projections routed to the auditory pathways of ferrets: receptive fields of visual neurons in primary auditory cortex. J Neurosci 12:3651–3664
Rogers SO (2012) Integrated molecular evolution. CRC Press, Boca Raton
Romanes GJ (1883) Mental evolution in animals. Kegan Paul, Trench and Co., London
Sartre J-P (1957) The transcendence of the ego (trans: Williams F, Kirkpstrtick R). Noonday, New York
Smith CUM (1982a) Evolution and the problem of mind: Part 1. Herbert Spencer. J Hist Biol 15:55–88
Smith CUM (1982b) Evolution and the problem of mind: Part II. John Hughlings Jackson. J Hist Biol 15:55–88
Smith CUM (1983a) Herbert Spencer’s epigenetic epistemology. Stud Hist Phil Sci 14:1–22
Smith CUM (1983b) Anatomical concepts and the problem of mind. J Social Biol Struct 6:381–392
Smith CUM (2005) All from fibres: Erasmus Darwin’s evolutionary psychobiology. In: Smith CUM, Arnott R (eds) The genius of Erasmus Darwin. Ashgate, Aldershot
Smith CUM (2007) Evolution and the problem of mind in the 19th century. In: Turnini SK (ed) Consciousousness and learning research. Nova Sciences Publishers, Inc., Hauppauge
Spencer H (1844a) A new view of the functions of intuition and benevolence. Zoist J Cereb Physiol Mesmerism 1:369–385
Spencer H (1844b) On the situation of the organ of attractiveness. Zoist J Cereb Physiol Mesmerism 2:186–188
Spencer H (1844c) A theory concerning the organ of wonder. Zoist J Cereb Physiol Mesmerism 2:316–325
Spencer H (1852) The development hypothesis. The Leader. Reprinted in Spencer 1878
Spencer H (1853) The universal postulate. Westminster Rev IV(NS) 4:513–550
Spencer H (1855) The principles of psychology. Longman, Brown. Green and Longman, London. Republished Gregg International, Farnborough
Spencer H (1864) Reasons for dissenting from the philosophy of M. Comte. Reprinted in 1878, Essays: scientific, political and speculative, 3rd edn. Williams and Norgate, London
Spencer H (1865) Mill versus Hamilton – the test of truth. Fortn Rev 1:531–550
Spencer H (1870a) First principles. Williams and Norgate, London
Spencer H (1870b) The principles of psychology, vol 1, 2nd edn. Williams and Norgate, London
Spencer H (1872) The principles of psychology, vol 2, 2nd edn. Williams and Norgate, London
Spencer H (1890) The principles of psychology, 3rd edn, 2 vols. Williams and Norgate, London

Spencer H (1904) An autobiography, 2 vols. Williams and Norgate, London
Von Baer KE (1828, 1837) Ueber Entwickelungsgeschichte der Thiere, 2 vols. Bornträger, Konigsberg
Wallace AR (1870) Natural selection and tropical nature. Macmillan, London. Reprinted 1969 Gregg International Publishers, Farnborough
Watson JB (1913) Psychology as the behaviourist views it. Psychol Rev 20:158–177
Webb B (1979) My apprenticeship, 2nd edn. Cambridge University Press, Cambridge
Wittgenstein L (1975) On certainty (ed: Anscombe GEM, von Wright GH). Blackwell, Oxford
Wittgenstein L (1976) Philosophical investigations (trans: Anscombe GEM). Blackwell, Oxford

Chapter 9
Problems of Consciousness in Nineteenth Century British and American Neurology

J. Wayne Lazar

9.1 Introduction

The writing of this chapter was motivated by David Chalmers' inquiry into problems of consciousness. Chalmers characterized problems of consciousness as "easy" and "hard."[1] Easy problems are explanations of cognitive functions, e.g., discrimination, integration of information, access to internal states, reportability of mental states, focus of attention, control of behavior, wakefulness, and sleep. These, according to Chalmers were satisfactorily explained by the end of the twentieth century or, at least, seemed to pose no unsolvable philosophical issues. Hard problems "seem resistant" to standard methods of cognitive science and explanations in terms of computational or neural mechanisms. The "really" hard problem is the explanation of the experience of consciousness: how can physical processes in the brain account for subjective experience.

The focus of this chapter is on problems of consciousness encountered by nineteenth century neurologists[2] when investigating instincts, reflexes, and localization of brain[3] functions. The discussion will show that consciousness was in the forefront

[1] Chalmers 1995.

[2] No single generic term accurately names the nineteenth century individual interested in neurophysiological explanations. For the most part, in this chapter the term will be neurologist, meaning a person interested in anatomy, physiology, and diseases of the nervous system.

[3] By way of orientation, instincts are typically not discussed anatomically. Reflexes are likely to have centers in the spinal cord, but the medulla and cranial nerves can be involved. The "reflex arc" clearly involves peripheral nerves and organs. Brain is a generic term whose meaning changes depending on its referent. The term "brain" in mind-brain interactions loosely refers to the "cerebrum," that upper portion above the brain stem. The centers for "brain functions," as in "localization of brain functions," are in cerebral cortices.

J.W. Lazar, Ph.D. (✉)
Independent Scholar, 155 Kensington Road South, Garden City South,
New York 11530, USA
e-mail: jwayne314@aol.com

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_9, © Springer Science+Business Media Dordrecht 2014

of neurophysiological inquiries and explanations. Unlike problems of consciousness in the twentieth century, in the nineteenth century they do not usefully sort into easy ones and hard ones. One was not more solvable than another. All problems of consciousness throughout the entire nineteenth century were difficult and all probably seemed intractable.

The mid-century explanation of voluntary activities[4] will illustrate just how important mental processes were in neurophysiological explanations. This explanation remained essentially unchanged throughout the century and for that reason it is called the standard model in this chapter. In 1836, Peter Roget (1779–1869), Fellow of the Royal Society, English physician, professor of physiology, and philologist, is quite explicit about a linear sequence of processes behind a voluntary act and the role of mental processes in this sequence:

> A voluntary action, occurring as the immediate consequence of the application of an external agent to an organ of the senses, though apparently a simple phenomenon, implies the occurrence of no less than twelve successive processes, as may be seen by the following enumeration. First, there is the modifying action of the organ of the sense, the refractions of the rays, for instance, in the case of the eye: secondly, the impression made on the extremity of the nerve: thirdly, the propagation of this impression along the nerve: fourthly, the impression or physical change in the sensorium. Next follow four kinds of mental processes, namely, sensation, perception, association, and volition. Then, again, there is another physical change taking place in the sensorium, immediately consequent on the mental act of volition: this is followed by the propagation of the impression downwards along the motor nerve; then an impression is made on the muscle; and, lastly, we obtain the contraction of the muscle, which is the object of the whole series of operations.[5]

According to Roget's summary, the first three and the last three processes are strictly physiological consisting of sensory organs, afferent and efferent nerve transmission, and motor organs. The middle four processes are mental processes consisting of sensations, perception, association, and volition. They, elements of the mind, are the origin and the basis of consciousness and intelligence. The remaining two processes—one on either side of the mental group—are mysterious processes. One converts afferent impulses into (mental) sensations and the other transforms mental processes back to nervous impressions. This standard model can be found in the writings of many important nineteenth century neurologists.[6] Principles of association [7] accounted for the accretion from more elementary mental elements to more complex ideas.

There was speculation about where in the nervous system the transformation from sensory impressions to sensations occurred and where further embellishments from sensations to perceptions to ideas took place. For many, this occurred in the

[4] The terms "activity" and "act" or their plurals were used instead of the current familiar term "behavior."

[5] Roget 1836, p. 376.

[6] Roget 1836; Abercrombie 1843; Carpenter 1845, 1875; Todd 1847; Paine 1840, 1849, 1872.

[7] The principles date back to Aristotle with additions by a string of British philosophers from John Locke (1632–1704) and Thomas Hobbes (1688–1779) to Bain. (Broadbent 1876; Bain 1883, 1894; Huxley 1890; Boring 1957; Young 1970.)

sensorium commune, usually thought to be in the brain, but its actual location was never fully discovered. From medieval times speculation about its location included the ventricles, the meninges, the surface of the cortex, the medulla, the central ganglia, or the hemispheres.[8] The Irish anatomist John Cleland (1835–1925) argued "the *sensorium* extends from the encephalon to the sensory origins".[9] For William James (1842–1910), the brain and mind are separate: cerebral processes have cells connected by fibers (this was the pre-neuron era) while mental processes have ideas connected by associations.[10] According to American physiologist Winfield Hall (1861–1942) the *sensorium*, wherever it is, implies that mental processing requires the man himself.[11]

For Alexander Bain (1818–1903), there is no "inner chamber" like the *sensorium' commune.*[12] "The organ of mind is not the brain by itself; it is the brain, nerves, muscles, and organs of sense" (p. 61). Bain provided one of the most influential definitions of the mind during the second half of the nineteenth century.[13] He defined mind as having three major elements: intelligence, feeling, and volition. According to Bain (1894) feeling "includes, but is not exhausted by, our pleasures and pains. Emotions, passion, affection, [and] sentiment, are names of Feeling…Thought is intellect or cognition. Volition "or Will, embrac[es] the whole of our activity as directed by our feeling" and intellect" (p. 2). Whether an animal feeds, fights, or flees, sensations and feelings furnish the stimulus or support for the activity. Consciousness links feeling (idea) to movement and "this feeling-prompted activity is called volition" (p. 4).

Even after continual discussions of mind and its importance in the explanation of volitional activities, its meaning in actual use is ambiguous in the nineteenth century. Perhaps the best way to demonstrate this is with Bain's use of the term. Ambiguity is evident in his use of the word "mental," mind's adjectival form, in his third edition of *Senses and Intellect* where "mental" modified at least 70 different nouns.[14] Thirty-three of these nouns appeared only once for example, "agitation," "cause," "disgust," "fatigue," "force," "outburst," and "suggestion." Only seven nouns appeared more than five times, namely, "state" or its plural, "science", "system," "excitement," "fact" or its plural, "action" (or actions or activity or activities), and "effects".

Commentators provided about the same level of insight at the latter part of the century as was available at mid-century. Even so, during this time, there was a move away from psycho-physical interactionism to psycho-physical parallelism. Language was more objective and commentators were more apologetic when

[8] Roget 1836.

[9] Cleland 1874, p. 111.

[10] James 1890.

[11] Hall 1905.

[12] Bain 1855.

[13] Bain 1855, 1864, 1868, 1894.

[14] These data were the result of searching the book with Google's find option. This number of instances of "mental" should only be understood to represent the vast majority of the actual number. The character recognition software used in this feature is not perfect.

alluding to mind-brain interactions. The American psychologist, George Ladd (1842–1921) described neurophysiological processes using mechanistic and physical terms, like “molecule” and “molecular motion.”[15] Although he insisted on describing mechanism, he necessarily brought mind into the process in a manner similar to Roget because at his time neurophysiology of the higher brain centers was still about mental functions. He was careful not to say that one was derived from the other, but he suggested a mutual agreement between them that emphasized the study of one is the study of the other. William James (1842–1910), American psychologist and philosopher, described the process in terms of reflexes and lower and higher brain centers. “Nerve-currents” from the sensory organs “arouse ideas in the hemispheres,” which direct and organize reflexes.[16]

By the end of the nineteenth century, researchers identified tropisms, reflexes, instincts, habits, intelligence behaviors, voluntary behaviors, non-voluntary behaviors, and even vital and complex behaviors. Each had specific characteristics and explanatory basis. These behaviors or “activities,” as they were called, were identified by a bewildering number of alternative and overlapping criteria, for example, whether the activities were simple or complex, whether instigating stimuli came from within or outside of the organism, whether the activities were performed with or without experience, whether the activities were performed habitually, that is, with or without attention, whether the activities are voluntary or involuntary, whether the activities were intended or not, and whether the activities were accompanied by consciousness or not.[17] A close look at attempts to define and explain instincts, reflexes, and localized brain functions will raise issues about consciousness and give insights into aspects of consciousness of most concern.

9.2 Instincts

Of course, Darwin just after mid-century provided an explanation for generating complex, adaptive behaviors that is strictly mechanical and based on the inheritance of “numerous, successive, slight modifications of simpler instincts.” With respect to bees making cells in their hives, no knowledge of “hexagonal prisms” or “basal rhombic plates” was required.[18] This explanation does not require knowledge of the adaptive significance of the current, simple instinct, that is, it does not require intent or consciousness. It does require transmission by inheritance of the successful instincts. Darwin’s approach eventually prevailed, but there were few with Darwin’s expertise who could follow his lead.[19]

---

[15] Ladd 1887.
[16] James 1890, p. 24.
[17] Romanes 1888; Morgan 1895.
[18] Darwin 1866, p. 280.
[19] Whitman 1899.

There were substantial discussions about the presence and absence of consciousness even by followers of Darwin. British psychologist and 'Darwinist' George Romanes (1848–1894) distinguished reflexes from both instincts and reason by the criterion of consciousness. British biologist and animal behaviorist C. Lloyd Morgan (1852–1936), was uncertain about the presence of consciousness. Writing about a simple versus complex dichotomy that he called vital and complex activities,[20] "All of these [both vital and complex] may be, and the last, the intelligent actions, invariably are, accompanied by consciousness".[21]

Both Romanes and Morgan are explicit about the difficulty in determining for sure whether consciousness accompanies an activity. Romanes wrote that imposing "a mental element" on instinct is controversial and it is "often difficult or even impossible, to decide whether or not a given action implies the presence of the mind-element—*i.e.,* conscious as distinguished from unconscious adaptation …."[22] Morgan identified consciousness objectively, by its classic signs. If the activity is adaptive, complex, or coordinated then mind is implicated in its origination. Yet, he stressed the difficulty deciding, on a case by case basis, the presence or absence of consciousness. For him, even "perfectly organized habitual activities are frequently in us unconscious".[23] Ultimately, the quest appeared to be whether the activity was intended.

For Bain and for Carpenter in the 1840s, as for James in 1890, intelligence[24] involves a means and an end and an understanding of both the means and the end. The deciding criterion is an implied feeling or an idea behind the response. The English physician and comparative physiologist William Carpenter (1813–1885), acknowledged that all adaptive behaviors were not necessarily voluntary and, thus, associated with intent. Esophageal contractions during swallowing are an example of an adaptive response that is not voluntarily.[25] Bain (1855) also acknowledged the similarity between non-voluntary and voluntary movements, associating the former with the spinal cord and the medulla oblongata and the latter with the cerebral hemispheres, corpora striata, and optic thalami. These types of movements use the same muscles and activate the same moveable parts, but non-voluntary movements do not require "feelings as an indispensable condition of their performance" and, thus are not considered under the province of mind.[26] Thus, the distinction between voluntary and involuntary movements—between intentional and unintentional activities or between conscious and unconscious activities—is metaphysical or psychological and not something directly observable, making its application to neurophysiological research tenuous.

---

[20] Complex activities include reflexes, instincts, and intelligent activities.

[21] Morgan 1895, p. 431.

[22] Romanes 1888, p. 11.

[23] Morgan 1895, p. 431.

[24] By intelligence, is meant these complex, adaptive, purposive, and etc. acts.

[25] Carpenter 1845.

[26] Bain 1855, p. 47.

9.3 Reflexes

Instinctive behaviors, such as hive making or nest building, can be exceedingly complex. So much so that they can stretch both knowledge and imagination to the fullest when considered neurophysiologically. Simpler responses, like Morgan's vital activities or Romanes' reflexes, proved more easily approachable. A simple stimulus may be followed by a simple response as if the stimulus is *reflected* back in the form of the response, thus the etymology of the word reflex as well as its meaning. An anatomical substrate of reflexes in the spinal cord had been suggested in the work of Bell, Magendie, Müller, and Hall.[27] Marshall Hall (1790–1857) was pivotal in establishing the reflex as "an explanatory principle in the interpretation of human and animal behavior" in the 1830s.[28] Hall[29] was adamant about restricting reflexes to the *medulla spinalis* (the spinal cord) and the *medulla oblongata*. His intent was to demonstrate a syndrome (of elicited movements, certain diseases, and poisonous effects) that he called reflex and that was independent of the brain, that is, the mind.

Similar questions about consciousness arose for reflexes as it did for instincts. Is consciousness involved in spinal reflexes even when the brain is severed from the cord? Despite the relative simplicity of a reflex, the movement is well coordinated and it seems to accomplish a particular goal. To some, its simplicity suggests a purely mechanical process; to others, its complexity appears by definition to require consciousness for its execution.

According to Fearing[30] there were differences of opinion about the presence of consciousness in reflexes among the predecessors and contemporaries of Hall. Predecessors, like Robert Whytt (1714–1766), Julian Legallois (1770–1814) and Herbert Mayo (1796–1852) thought some form of consciousness was available to the cord. While other predecessors like Johann Unzer (1727–1799) and Gilbert Blane (1739–1834) and other contemporaries like Jean Flourens (1794–1867), Louis-Florentin Calmeil (1798–1895), Richard Grainger (1801–1865), and Carpenter did not find evidence of consciousness in the cord. The list can be expanded to include Alfred Volksmann (1800–1877), Eduard Pflüger's (1829–1910), Rudolph Lotze (1817–1881) and, later, William Hammond (1828–1900) and Michael Foster (1836–1907) who thought consciousness was available in the cord.

The German physiologist Volkmann, a contemporary of Hall and Carpenter, was one of several researchers who questioned Hall's mechanical theory. His ideas were particularly important, because of their influence on subsequent discussions that extended well past mid century. Volkmann was impressed that decapitated frogs responded in a coordinated and goal-directed fashion. From this, Volkmann concludes, "that the decapitated animal is aware of the action of

[27] Hodge 1890a; Fearing 1930.

[28] *Ibid.*, 1930, p. 145.

[29] Hall 1833.

[30] *Ibid.*, 1930.

the stimulus, and chooses, amongst a variety of means, those which are best calculated to free itself from the annoyance".[31] His conclusion did not appear to emphasize the experience of consciousness as much as it emphasized that conscious processes accounted for the frog's activities. Volkmann was not impressed with Hall's reasoning, which was based on the observation that the decapitated frog lacked spontaneity—that it was acting without cooperation of the mind. Volkmann saw many movements that he attributed to discomfort and thus mind. For Volkmann, there is no tenable ground for denying the cooperation of a psychic principle although his qualification was that its action must be blunted when functioning in the spinal cord alone.[32] "Sensations are certainly perceived, only they must be more obtuse and very limited after the removal of the specific organ of sensation [, the brain]".[33]

Volkmann argued for psychical processes in reflex action and many of his objections to Hall's theory influenced the German physiologist Pflüger's critique of unconscious reflexes in 1853.[34] Pflüger's critique began a controversy with Lotze that attracted a great deal of discussion. For Pflüger consciousness is a function of nervous action. It occurs everywhere in the nervous system and cannot be excluded from the spinal cord even when it is divided from the brain.[35] He accepted purposive and well coordinated movements of reflexes as evidence for his point of view. His position is consistent with Cartesian unity of mind. Lotze, on the other hand, emphasized that consciousness depended on the brain and would not be found in the spinal cord when isolated from the brain. The spinal animal was mechanistic, but it was not mechanical when intact, because mind always takes part in behavior of an intact organism. The mind worked through reflexes to construct behavior. The will chooses the function, but not the route.

The controversy remained unresolved and disagreements occurred throughout the century. Even if there was agreement that consciousness existed in the spinal animal, the reason for its existence may have been moot. For example, William Hammond (1828–1900), an exceedingly well respected American neurologist and one with a strong materialist bent,[36] defended the conclusion of consciousness in the spinal cord based on the purposive behavior of the decapitated frog and the assumption that where there is gray matter there is mind.[37] The eminent, English physiologist Michael Foster (1836–1907), although not impressed with the decapitated frog's purposive behavior, was impressed by the complexity of its behavior, which, for him, indicated of the presence of mind. Such complexity, such coordination required a mind, consciousness. Here we clearly have the presence of consciousness meaning more than just the experience of it. Consciousness had a

---

[31] Anonymous reviewer, 1838, p. 213.

[32] Hodge 1890b.

[33] *Ibid*., p. 350.

[34] Fearing 1930.

[35] Gault 1904; Boring 1957.

[36] Blustein 1991.

[37] Hammond 1876.

function, namely, to account for the complexity. Foster went so far as to postulate a momentary or intermittent consciousness in the spinal animal. "We may thus infer that when the brainless frog is stirred by some stimulus to a reflex act, the spinal cord is lit up by a momentary flash of consciousness coming out of darkness and dying away into darkness again".[38]

9.4 Localization of Brain Functions

Localization of brain functions is a milestone in nineteenth century neurophysiology. Localized brain functions occurred in brain centers, which are ganglia in the cerebral cortex. The cells of each ganglion were thought to act together in some unknown way and to be responsible for the function. Kinds of functions include those for language (sensory and motor), motor (for movement of various muscles), and sensory (for seeing, hearing, touch, temperature, and muscle sense, i.e., proprioception). Research effort in the latter part of the nineteenth century attempted discover these functions and delimit their specific cortical locations. Study of these research efforts show the same concerns about the presence of mind and consciousness raised in conjunction with instincts and reflexes plus concerns about the isomorphism between brain functions and mental functions.

Nineteenth century localization of mental functions in the brain began with Franz Gall (1758–1828),[39] who attained great prominence with his localization system, "organology," which was popularized as "phrenology" through lectures and books by Spurzheim and Combe. Gall's impact derived from his emphasis on functional aspects of the brain and the concrete nature of his faculties. His organology described personality characteristics based on the localization of mental faculties.

Gall postulated that the cerebrum is composed of different functional regions each associated with a different moral or intellectual faculty. These faculties are innate and "their exercise or manifestation depends on organization" of the brain.[40] The size of a region directly correlates with a person's personality. Surprisingly, for a respected anatomist, Gall did not study brains when locating brain centers for his faculties; he studied skull topography, inferring the location of functional brain regions from prominent regions of the skull on the assumption that the size of a region directly affects the shape of the skull. Phrenological theory contradicted the Cartesian view of an indivisible mind.

Of particular relevance to acceptance of localization, is Descartes' insistence that mind is unitary and indivisible. Some 200 years earlier, Descartes had written, "There is a great difference between the mind and the body, in that the body, by nature, is always divisible, and the mind is entirely indivisible".[41] Although the

[38] Foster 1890, p. 912.

[39] Young 1970.

[40] Gall 1835.

[41] Morris 1971, p. 136.

mind is associated with the brain, the mind remains unitary when the brain is partitioned much like a fractal remains whole after analysis. “I do not now conceive that the mind is extended in the body otherwise [than weight], when I conceive it to be completely contained in the whole, and to be completely contained in each part”.[42]

Although Gall’s views captured wide attention in popular culture, unity of mind remained dominant in the French Academy and in neurology during the first half of the nineteenth century.[43] The French physiologist, Jean Pierre Flourens (1794–1867) led the anti-phrenological movement. Although Flourens’ conclusions about phrenology were reasonable interpretations of his own experimental results with animals, something besides cold, clean data drove Flourens. After all, he dedicated his book to Descartes. He, his sponsor in the Academy, Georges Cuvier, and the majority of the French academy were fundamentally against multiple organs of intelligence.[44]

Major influences were needed to overturn this tradition. Opinions like those of Flourens[45] dominated through about the 1850s, but weakened afterward. Among these influences were a conceptual change in the meaning of language, clinical results in the form of autopsies of aphasics, and experimental results in the form of brain stimulation in mammals and primates. Paul Broca (1824–1880), French physician and anatomist, published two case studies of autopsied patients with aphasia in 1861, implicating the left frontal lobe; he followed these with a more specific report in 1865 that located the lesion in the third, left frontal convolution.[46] Autopsy evidence for language localization in the frontal lobe had been available in the past: Jean-Baptiste Bouillaud (1796–1881) had implicated the frontal lobes by 1825. However, as a supporter of Gall, his clinical cases were ignored.[47] In contrast, Broca’s announcement was associated with a lively debate about localization and a naturalistic approach to brain functions during meetings of Société d’Anthropologie, which he founded. Language was becoming a bodily function rather than a vehicle for reasoning and thought. More generally, metaphysics had lost ground. “Metaphysics might have its place; but it had no use in understanding man conceived as a distinct, observable and measureable [*sic*] object”.[48]

In 1870, Gustave Fritsch (1838–1927) and Eduard Hitzig (1838–1907) and, in 1873, David Ferrier (1843–1928) published evidence that specific movements

[42] Morris 1971, p. 137; The bracketed information is in the original quote.

[43] Flourens 1846.

[44] Boring 1957; Finger 2000

[45] Not only his experimental evidence convincing, but Flourens expressed concern that Gall’s theories would undermine free will, human immortality, and the very existence of God. Harrington 1989. It suggested materialism. These ideas were pervasive in Catholic-Europe. *The Phrenological Journal and Miscellany* of 1830 reported that Professor Uccelli of the University of Pisa was dismissed from his long held chair after he advocated one of the volumes of Gall and Spurzheim’s “Physiology of the Brain,” which was held as synonymous with fatalism, materialism, atheism, and, worse than all, Protestantism.

[46] Broca 1960; Finger 1994.

[47] Jacyna 2000.

[48] *Ibid*., p. 77.

were associated with electrical stimulation of specific brain areas along the Rolandic fissure.[49]

Fritsch and Hitzig's demonstration of motor functions in unique areas of the brain of dogs may have come as a surprise to some, but not to John Hughlings-Jackson (1835–1911), who had postulated localization of motor and perceptual functions in the 1860s based on his clinical observations; he recognized that a somatotopic organization of the cortex was the only way to account for the progression of seizures.[50] Vindicated by the results of Fritsch and Hitzig, Huglings-Jackson encouraged Ferrier to follow up the localization effort. Not only did Ferrier replicate and extend Fritsch and Hitzig's results, but also he popularized it[51]; localization became an international phenomenon.

Explaining such functions would be among the "easy" problems of Chalmers, but they were not easy for neurologists of the time. Disagreements ensued about their very existence. Initially many neurologists were leery about these elicited brain functions. Eventually, most neurologists came around to agreement on some form of their existence. Some physiologists, like Charles Brown-Séquard (1817–1894), continued to argue against localization interpretations. He promoted the traditional interpretation that cells associated with specific movements were "scattered" throughout the hemispheres such that there were no clusters of cells that could be called motor centers.[52] He and his student, Eugene Dupuy, opted for this Cartesian-friendly view.[53]

Others like Hughlings-Jackson and Wilhelm Wundt (1832–1920) accepted localized brain functions, but disputed their psychological interpretation. Influenced by Herbert Spencer (1820–1903) and Bain, Hughlings-Jackson advocated a strict psychophysical parallel approach to neurology in order to avoid the pervasive tendency to imply an interaction between mental processes and the nervous system.[54] Psychophysical parallelism leaves neurologists free to investigate the "substrata of consciousness" or "substrata of mind" in the brain, sensori-motor processes—only impressions and movements—without the need for interactions or explanations.[55] Wundt emphasized a similar approach. He argued that neurologists work with the nervous system, not complex psychological phenomena. "'Everything, that we call will and intelligence resolves itself, as soon as it is traced back to its physiological elements, into nothing but sentient impressions transforming themselves into movements'".[56] In another place, Wundt writes about language localization, "We cannot possibly imagine, from what we know either of the brain or of the psychical processes, that a definitely circumscribed brain area is the seat of linguistic endow-

[49] Fritsch and Hitzig 1963; Ferrier 1873a, b; Finger 1994.

[50] Finger 2009.

[51] Lazar 2009a.

[52] Brown-Séquard 1878a, b, c; Jewell 1878.

[53] Lazar 2009b.

[54] Young 1970; James 1890.

[55] Jackson 1875, p. xxiii, xxx.

[56] Lange 1881, p. 154.

ment, in the same sort of way that eye and ear are organs of the reception of light and sound stimuli".[57] His introspection analyzed mental processes, but his physiological approach did not.

Carl Wernicke (1848–1905) accepted individual functions, broadened the explanation of aphasia, and continued neurophysiology squarely within the traditional model associated with mental processes. He argued that aphasia, like other mental functions, is a symptom complex composed of primitive "memories" of past sensory and motor experiences, which are associated and combined according to an associationist-connectionist brain model.[58] Wernicke's emphasis on "system complex" promoted system analysis rather than specific functional centers that accounted for the entire process.

Regardless of comments like those from Hughlings-Jackson and Wundt, there was a quest for unique brain centers for mental function. Mind is associated with all intelligent behavior and is the originator of the intelligent behavior. Whether motor centers, sensory centers, or centers of complex functions like language, researchers behaved as though they expected to find centers in particular spots in the brain. When a lesion of an area resulted in elimination of a function, the assumption was that the area was the center for that function. If a patient or animal subject recovered the function, as the latter often did, then the center had not been destroyed. In the latter case, according to the point of view of their sequential model of the nervous system, only an associated function in the path had been affected. To make a railroad analogy, they had not eliminated the terminal, just a station on the line. The German journalist Frederick Lange (1828–1875) makes these arguments in his book the History of Materialism. Lange concluded that this type of thinking was typical.[59]

Some neurologists, like James Jewell (1837–1887), Ferrier, and John Dalton (1824–1889) found a way to maintain a role for hemispheric motor areas, to acknowledge recovery of them, and yet stay within the tradition of the standard model for explaining voluntary activities. Jewell defended the role of centers in the cortex in spite of recovery, reasoning that centers in the hemispheres enabled the will to influence the "true motor system below".[60] The cerebral centers did not originate functions, just allowed and directed them. The true motor centers, such as those from non-cortical areas like the basal ganglion, were excitable from multiple sites so it was not surprising to Jewell that a motor function would return even if volition was not in control. Ferrier (1876) and Dalton (1882) came quite close to postulating mechanistic motor centers without relinquishing the role of mind; their position seems intermediate between a nervous system necessitating mind and a mechanistic nervous system not requiring mind. They made an argument similar to Jewell's, but using species differences and a Darwinian approach to account for recovery.

[57] Wundt 1904, p. 296.

[58] Harrington 1989.

[59] Lange 1881.

[60] Jewell and Bannister 1877, p. 557.

For them, recovery is a species effect that depends on the role of the cortex[61] and volition in the animal's nervous system. Top-down control is limited in amphibians and birds so brain lesions barely affect them and they recover rapidly and fully. Lesions affect "lower" mammals more and the effects last longer. Cortical lesions in primates, including humans leave them essentially permanently paralyzed. These differences occur because top-down control plays a greater role in volition (cortical control) in the more complex animals. "[I]n the higher animals, and especially in man, the influence of immediate volitional impulses is more essential, and preponderates in importance, according to the number and variety of the muscular actions".[62]

9.5 Summary and Conclusions

The foregoing analysis shows that nineteenth century neurologists studied the neurophysiological bases of several types of activities, including reflexes, instincts, and voluntary activities. The sequence of processes that account for voluntary, namely, intelligent, activities (illustrated by Roget in 1836 and called in this chapter, "the standard model") persisted throughout the century and included central processes of psychic origin. Neurologists agreed the transmission of the nerve impulse was purely mechanical, "but [not] the central process by which a sensory is changed to a motor impulse, and so directed as to cause definite movements of the muscles?"[63] There were continuous debates about mental versus mechanical solutions for the central process, that is, about the presence of mind in various activities. The psychic solutions were elaborate schemes about mental processes, such as, those in the standard model and those discussed by Bain. The mechanical solutions were not elaborate. They were of the kind, "ganglia do it" or "psychic processes cannot be in the nervous system." There is a variety of reasons for the paucity and simplicity of mechanical solutions including neurophysiological, traditional, and religious ones that are beyond the scope of this chapter.[64] From the nineteenth century, neurophysiological point of view, a satisfactory mechanism was not conceivable.

Intelligent activities were identified by their complexity, coordination, purposiveness, and adaptive value. Problems arose when applying these criteria to other kinds of activities. Some adaptive activities, for example, esophageal contractions, were easily accepted to be outside of conscious control, but what of activities like reflexes and instincts and what about functions localized by electrical stimulation? Analyses in this chapter identified three neurological problems related to this quest. One is about whether consciousness is present during activities, such as, reflexes

[61] Ferrier 1876. Since Ferrier included amphibians and birds as well as mammals in his discussion, I will use the term top-down control to be more anatomically correct.

[62] Dalton 1882, p. 429.

[63] Hodge 1890b, pp. 350–351.

[64] Lazar 2012.

and instincts. A second is about whether the mind is present as a Cartesian-like unity or as multiple and interacting processes scattered throughout the brain and the interpretation of functional brain centers. The third is about the relationship of the mind and the brain, for example, is the relationship dualistic or monistic.

None of these problems was resolved during the century. Identifying behavior as coordinated, complex, purposeful, and adaptive proved ambiguous. Bain's identification of intelligent behavior with a "feeling" or a knowledge of means and end takes it outside the bounds of neurophysiology. The presence of mind did not easily translated into physiological terms, yet it was part of the standard neurophysiological model and the only way, at that point, to explain intelligent behavior. There was no resolution throughout the entire century about whether instincts included consciousness. A mechanical explanation for reflexes, at least spinal reflexes, was easier to accept. Sherrington's is the best of this type by the end of the century.[65]

Unity of mind predominated for about the first half of the century and directed expectations about finding relatively independent cerebral functions. Ideas of multiple, relatively independent brain functions dominated during the last quarter of the century. Since brain functions were considered centers for mental processes, the idea of unity of mind was less appealing. Nevertheless, fundamentally this problem remained unresolved because the characterization of interacting multiple parts remained as mysterious as was the characterization of the functional whole.

The mind-brain relationship for neurologists remained predominantly dualistic, but moved from a metaphysical relationship in which the two interacted in some mysterious fashion to an agnostic-like position that allowed both to be studied without concern about the other while knowing, full well, that they were related somehow.

The analysis of problems related to consciousness in nineteenth century neurophysiology did not reveal a quest to understand the experience of consciousness. The analysis was about the presence of consciousness in various activities and the interpretation of the neurophysiological bases of what were to become cognitive processes. The latter would be identified later as one of Chalmers' easy problems. The outstanding neurophysiological problems related to consciousness of the nineteenth century were either solved or solvable, according to Chalmers, by the end of the twentieth century. Chalmers' hard problem revealed itself in the late twentieth century phenomenon. Perhaps, like with nineteenth century difficult problems for which no solution was conceivable at the time, Chalmers' hard problem of the twentieth century will promote discussions, disagreements, and alternative suggestions for the next number of years. Such so-called impenetrable problems, certainly seems less unusual and less mysterious after viewing the tribulations of nineteenth century neurologist. Maybe, as Dennett suggests, it will just go away[66] or maybe it will be classified as "easy."

---

[65] Sherrington 1906.

[66] Dennett 1991.

References

Abercrombie J (1843) Inquiries concerning the intellectual powers, and the investigation of truth. Otis, Broaders, and Company, Boston
Bain A (1855) The senses and the intellect. John W Parker and Son, London
Bain A (1864) The senses and the intellect. Longman, Green, Longman, Roberts, and Green, London
Bain A (1868) The senses and the intellect. Longmans, Green, and Co, London
Bain A (1883) Mind and body. Mind 8:402–412
Bain A (1894) The senses and the intellect. D Appleton and Company, New York
Blustein BE (1991) Preserve your love for science. Cambridge University Press, Cambridge
Boring EG (1957) A history of experimental psychology. Appleton-Century-Crofts, Inc., New York
Broadbent WH (1876) A lecture on the theory of construction of the nervous system, part 3. Br Med J 1:433–436
Broca P (1960) Remarks on the seat of the faculty of articulate language, followed by an observation of aphemia. In: von Bonin G (ed) The cerebral cortex. Charles C. Thomas, Springfield, pp 49–72
Brown-Séquard CB (1878a) Lecture 1, paralysis and convulsions as effects of diseases of the base of the brain. Chic Med J Exam 36:291–305
Brown-Séquard CB (1878b) Lecture 2, paralysis and convulsions as effects of diseases of the base of the brain. Chic Med J Exam 36:353–366
Brown-Séquard CB (1878c) Lecture 3, paralysis and convulsions as effects of diseases of the base of the brain. Chic Med J Exam 36:449–465
Carpenter WB (1845) Principles of human physiology. Lea and Blanchard, Philadelphia
Carpenter WB (1875) Principles of mental physiology. D. Appleton and Company, New York
Chalmers DJ (1995) Facing up to the problem of consciousness. J Conscious Stud 2(3):200–219
Cleland J (1874) Animal physiology. William Collins Sons, & Company, London/Glascow
Dalton JC (1882) A treatise on human physiology. Henry C. Lea's Son & Co, Philadelphia
Darwin CR (1866) Origin of the species. John Murray, London
Dennett DC (1991) Consciousness explained. Little, Brown, Boston
Fearing F (1930) A study in the history of physiological psychology. The Williams and Wilkins company, Baltimore
Ferrier D (1873a) Experimental researches in cerebral physiology and pathology. J Anat Physiol 8:152–155
Ferrier D (1873b) Experimental researches in cerebral physiology and pathology. West Riding Lunatic Asylum Report, vol 3. London, pp 30–96
Ferrier D (1876) The functions of the brain. Smith, Elder, London
Finger S (1994) Origins of neuroscience. University Press, Oxford
Finger S (2000) Minds behind the brain. University Press, Oxford
Finger S (2009) The birth of localization theory. In: Finger S, Boller F, Tyler KL (eds) Handbook of clinical neurology, vol 95, (3rd series). Elsevier, pp 117–128
Flourens P (1846) Phrenology examined, 2nd edn. Hogan & Thompson, Philadelphia
Foster M (1890) A text book of physiology, part 3. Macmillan and Co, London
Fritsch G, Hitzig E (1963) The electrical excitability of the cerebrum, in neurosurgical classic XII (trans: Wilkins RH). J Neurosur 20: 904–916
Gall FJ (1835) On the functions of the brain, vol 1. Marsh, Capen, & Lyon, Boston
Gault RH (1904) A sketch of the history of reflex action in the latter half of the nineteenth century. Am J Psychol 15:526–568
Hall M (1833) On the reflex function of the medulla oblongata and the medulla spinalis. Philos Trans R Soc 123:645–665
Hall SW (1905) A text-book of physiology. Lea Brothers & Co, Philadelphia
Hammond WA (1876) The brain is not the sole organ of the mind. J Nerv Ment Dis 3:1–19

Harrington A (1989) Beyond phrenology: localization theory in the modern era. In: Corsi P (ed) The mill of thought. Elemond Editorial Association, Milano, pp 206–239
Hodge CF (1890a) A sketch of the history of reflex action, part 1. Am J Psychol 3(2):149–167
Hodge CF (1890b) A sketch of the history of reflex action, part 2. Am J Psychol 3(3):343–363
Huxley TH (1890) Lessons in elementary physiology. Macmillan and Company, London
Jackson J (1875) Clinical and physiological researches on the nervous system reprints. J and A Churchill, London
Jacyna LS (2000) Lost words: narratives of language and the brain 1825–1926. Princeton University Press, Princeton
James W (1890) The principles of psychology, volume I. Henry Holt and Company, New York
Jewell JS (1878) On Dr. Brown-Séquard and his recent lectures. Chic Med J Exam 36:561–571
Jewell JS, Bannister HM (1877) American Neurological Association Meeting, June 6–8, 1877, second day's proceedings. J Nervous Mental Dis 4*:* 554–557
Ladd GT (1887) Elements of physiological psychology. Charles Scribner's Sons, New York
Lange FA (1881) Lange's history of materialism, vol 3. Houghton, Mifflin & Company, Boston
Lazar JW (2009a) Anglo-American interest in cerebral physiology. J Hist Neurosci 18:304–311
Lazar JW (2009b) Diffusion of electrical current in the experiments of Fritsch and Hitzig and Ferrier failed to negate their conclusion of the existence of cerebral motor centers. J Hist Neurosci 18:366–376
Lazar JW (2012) Brain physiology and the mind in the nineteenth century. J Hist Neurosci 21:343–365
Morgan CL (1895) Animal life and intelligence. Ginn & Company, Publishers, Boston
Morris J (1971) Descartes dictionary. Philosophical Library, New York
Paine M (1840) Medical and physiological commentaries, vol 2. Collins, Keese & Co, New York
Paine M (1849) A discourse on the soul and instinct physiologically distinguished from materialism. Edward H Fletcher, New York
Paine M (1872) Physiology of the soul and instinct as distinguished from materialism. Harper & Brothers, Publishers, New York
Roget PM (1836) Animal and vegetable physiology considered with reference to natural theology, vol 2. Carey, Lea & Blanchard, Philadelphia
Romanes GJ (1888) Instinct. In: Encyclopedia Britannica, vol 13. Henry G. Allan and Company, New York, pp 157–160
Sherrington CS (1906) Integrative action of the nervous system. Yale University Press, New Haven
Todd RB (1847) Cyclopaedia of anatomy and physiology. Longman, Brown, Green, and Longmans, London
Whitman CO (1899) Animal behavior. Biological lectures of the marine biological laboratory, Wood's Holl, MA, pp 285–338
Wundt WM (1904) Principles of physiological psychology, vol 1. San Sonnenschein & Co. Lim, London
Young RM (1970) Mind, brain and adaptation in the nineteenth century. Clarendon, Oxford

Chapter 10
Emil du Bois-Reymond's Reflections on Consciousness

Gabriel Finkelstein

10.1 The Limits of Science

> You are right in demanding that an artist approach his work consciously, but you are confusing two concepts: *the solution of a problem and the correct formulation of a problem.* Only the second is required of the artist.
>
> Anton Chekhov to Alexei Suvorin, 27 October 1888

Emil Heinrich du Bois-Reymond (1818–1896) grew up divided. His father was Swiss and poor, but his mother came from one of Berlin's most celebrated families. At home he learned art, music, and conversation, but his parents encouraged exercise, exploration, and experiment. A lifelong enthusiast of literature, history, and philosophy, he trained in geology, physics, and physiology. He enjoyed the respect of colleagues at the University of Berlin and the Prussian Academy of Sciences, but most of his friends were Jews and foreigners. During the term he worked in Berlin, but on weekends he commuted to Potsdam where he spoke German to his family, English to his servants, and French to his guests. Even his house was diverse: an Italian villa designed by a Prussian architect, it showcased Roman statues, Dutch paintings, German furniture, French books, and English vegetables. Such a cosmopolitan existence allowed him to see farther than most.

Du Bois-Reymond's perspicacity informed "The Limits of Science", a keynote address that he delivered to the Congress of German Scientists and Physicians in Leipzig on 14 August 1872. In this, the most famous of his speeches, du Bois-Reymond aimed to show that even the empire of science could not expand indefinitely.[1] To do so he first defined what he meant by natural knowledge. The reduction

[1] Du Bois-Reymond 1912d, p. 441. My translations are based on du Bois-Reymond, E. 1874 and my discussion is based on Finkelstein 2013.

G. Finkelstein (✉)
History Department, University of Colorado Denver, Denver, Colorado, USA
e-mail: Gabriel.Finkelstein@ucdenver.edu

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_10,

of events to the "apodictic certainty" of physics offered the only form of understanding that satisfied our desire for causal explanation.[2] The task of science, as du Bois-Reymond described it, was to construct a mathematical model of reality.

The astronomer Pierre-Simon Laplace first spelled out the implications of this project. "A mind which at a given instant should know all the forces acting in nature, as also the respective situation of the beings of which it consists—provided its powers were sufficiently vast to analyze all these data—could embrace in one formula the movements of the largest bodies in the universe as well as those of the smallest atom; nothing would be uncertain for such a mind, and the future, like the past, would be present to its eyes."[3] It could tell us, as du Bois-Reymond envisioned, "the day when the Greek cross shall glitter from the mosque of St. Sophia, or when England shall have burnt the last of her coals", or alternatively, "who was the Man in the Iron Mask, or how the *President* was lost".[4] To such a mind all things would become "one single fact and one great truth".[5]

Of course, du Bois-Reymond never expected science to reach this degree of perfection. To do so presumed that we could resolve natural events into the "vibrations of a primitive, undifferentiated matter"—a transmutation still awaiting the philosopher's stone.[6] There was also the practical impossibility of gathering all the necessary facts and tracing their infinite ramifications. Nevertheless, our ignorance was more a question of degree than of kind. If we did know the disposition of every atom, we could indeed calculate the fate of the universe with the confidence expressed by Laplace. Such an "astronomical intelligence" represented the fullest possible understanding of the world, since any information barred to it would necessarily be foreclosed to us.[7]

Having laid the foundation of his argument, du Bois-Reymond identified two bounds of understanding that even his "Laplacian demon" could not cross. The first was the essence of matter. Material atoms supplied a "useful fiction" for many considerations of physics, but volume-elements made more sense of continuous fields.[8] Metaphysical atoms presented similar difficulties. On the one hand, they were presumed to be inert points; on the other, they occupied space and interacted with the world.[9] What was worse, these contradictions appeared inevitable. Metaphysical atoms did little more than import into "the minute and the invisible

---

[2] Ibid., 1912d, pp. 442–443.

[3] Laplace 1814, pp. 3–4.

[4] du Bois-Reymond 1912d, p. 443. The Hagia Sophia was an Orthodox patriarchal basilica until 1453. "Lestang", The Man in the Iron Mask, was imprisoned by Louis XIV. His identity was never disclosed. The British passenger liner SS *President*, the largest ship in the world, left New York for Liverpool on 11 March 1841 and vanished without a trace.

[5] Ibid., 1912d, pp. 443–444, quoting d'Alembert 1893, p. 48.

[6] Of quantum mechanics, it turned out. Ibid., pp. 445–446.

[7] Ibid., p. 446.

[8] Ibid., p. 447.

[9] Ibid., p. 448. How they did this was a mystery: action at a distance contradicted itself, and action across media failed to explain variations in density.

the qualities of the gross and the visible"; the same went for material atoms, which could not be expected to develop novel properties simply because they were small.[10] On either conception matter remained a riddle.

The origin of life might appear to set the other limit, but du Bois-Reymond assured his audience that this problem did not amount to an impenetrable mystery. The Laplacian demon could determine the precise conditions that produced the first living things. Organic matter was composed of common atoms; the puzzle was how they remained in dynamic equilibrium. Since the supernatural was excluded, biology translated to an "exceedingly difficult mechanical problem" that science might someday expect to solve. "For the rest", du Bois-Reymond wrote, "the most luxuriant picture of a jungle ever sketched by Bernardin de St. Pierre, Alexander von Humboldt, or Eduard Pöppig offers to the view of theoretical science nothing more than matter in motion".[11]

Instead, the second limit was consciousness. Having arisen at some point in the evolution of life, it was the one aspect of nature that could not be reduced to a material substrate. This held as true for plain sensations as for complex ideas: the "first awakenings of pleasure or pain in simple organisms" confronted the world with "an impassable gulf that rendered it doubly incomprehensible".[12] As du Bois-Reymond recounted, no philosophical reflection had come close to explaining how the mind interacted with the body—not Descartes' invocation of the pineal gland, nor Malebranche's appeal to divine assistance, nor Leibniz's assumption of perfect harmony. The best one could do was to regard consciousness as an effect of matter.[13]

With this du Bois-Reymond arrived at the heart of his argument. It certainly would be a "lofty triumph" if science could correlate mental phenomena with physiological activity. We could note with interest "what play of carbon, hydrogen, nitrogen, oxygen, and phosphorus corresponds to the bliss of hearing music, what whirl of such atoms answers to the climax of sensual enjoyment, what molecular storm coincides with the raging pain of trigeminal neuralgia." But even perfect knowledge of the brain would tell us nothing about experience, for "no imaginable movement of material particles could ever transport us into the realm of consciousness".[14] The same could be said of attempts to address the question psychologically, since perception, association, and memory could never substitute for awareness. In a famous passage, du Bois-Reymond laid out the difficulty: "What conceivable connection exists between definite movements of definite atoms in my brain on the one hand,

[10] Ibid., p. 449.

[11] Ibid., pp. 451–452. Jacques-Henri Bernardin de St. Pierre (1737–1814) was a writer and botanist best known for his novel *Paul and Virginia*. Eduard Pöppig (1798–1868) was a scientific explorer of South America. Du Bois-Reymond preferred the soberer language of Darwin and Moltke. Emil du Bois-Reymond to Jeannette Claude, 9 July 1853, Staatsbibliothek Preußischer Kulturbesitz zu Berlin, Haus 2, Handschriftenabteilung, hereafter SBBPK, Dep. 5 K. 11 Nr. 5; du Bois-Reymond 1912f, p. 57.

[12] Ibid., p. 453.

[13] Ibid., pp. 453–455.

[14] Ibid., p. 457.

and on the other hand such primordial, indefinable, undeniable facts as these: *I feel pain or pleasure; I taste something sweet, or smell a rose, or hear an organ, or see something red*, and the certainty that immediately follows: *Therefore I am*?" Even if the atoms of the brain were mindful of their own existence, science would be at a loss to explain how consciousness followed from their combined action.[15] "In a world made up of matter in motion", he declared, "the movements of the cerebral molecules are like a dumb show".[16]

At this point in his address du Bois-Reymond advanced a pair of theological considerations. The first was that the unconscious mind held no secrets for the Laplacian demon—the brain of a "dreamless sleeper" was as intelligible to it as were the orbits of the planets. This did not mean that identity was rooted in the will; sensation clearly preceded desire, a fact that implied that sin was subordinate to perception.[17] This led du Bois-Reymond to his second point: religion held no authority over science. Our knowledge was indeed "imprisoned by two limits", but between these "the man of science is lord and master; he can analyze and synthesize, and no one can fathom the extent of his knowledge and power." Science therefore could safely ignore "myth, dogma, or time-honored philosophy".[18]

As du Bois-Reymond saw it, the mind depended entirely on the brain. Ideas derived from the senses, morbid states altered thought, and animals experienced the world, all of which indicated that intelligence had emerged as a consequence of natural selection. In this regard the scholastic presumption that separated mental phenomena from material conditions was "so plainly in conflict with reality" that it supplied an "apagogical demonstration of the falsity of its premises".[19] The mystery of consciousness could not excuse the error of dualism.

Du Bois-Reymond brought his address to a surprising conclusion. Leibniz once conceived a superior intelligence "constructing a body capable of mimicking the actions of person." However, he considered this automaton to lack the "monad of the soul." Du Bois-Reymond pictured the *Doppelgänger* more classically:

> Imagine all the atoms of which Caesar consisted at any given moment, say, as he stood at the Rubicon, to be brought together by mechanical artistry, each in its own place and possessed of its own velocity in its proper direction. In our view Caesar would then be restored mentally as well as bodily. This artificial Caesar would have the same sensations, ambitions, and ideas as his prototype on the Rubicon, and would share the same memories, inherited and acquired abilities, and so forth.

"Suppose several artificial figures of the same model", he continued, "to be simultaneously formed out of a like number of other atoms of carbon, hydrogen, etc. What would be the difference between the new Caesar and his duplicate, beyond the

---

15 Ibid., p. 458.

16 Ibid., p. 460.

17 Ibid., p. 459.

18 Ibid., pp. 460–461.

19 Ibid., pp. 461–462. Cf. J. S. Mill 1859: "… no rational person can doubt the closeness of the connexion between functions of the nervous system and the phenomena of mind, nor can think any exposition of the mind satisfactory, into which that connexion does not enter as a prominent feature (p. 295)."

differences in the places where they were formed? But the mind imagined by Leibniz, after fashioning the new Caesar and his many Sosiae, could never understand how the atoms he had arranged and set into motion could lead to consciousness".[20]

Du Bois-Reymond recalled Carl Vogt's assertion that "thought is to the brain what bile is to the liver or urine is to the kidneys."[21] The weakness of the comparison was less aesthetic than intellectual: it suggested that consciousness could be explained by the structure of the nervous system in the same way that secretion could be explained by the structure of a gland.[22] Monism might offer the most practical philosophy of science, du Bois-Reymond conceded, but "whether we shall ever understand mental phenomena from their material conditions is a very different question from whether these phenomena are the product of their material conditions".[23] He doubted the prospect of a solution. Scientists were used to admitting their ignorance, "but as regards the enigma of matter and force, and how they are capable of thought, we must resign ourselves once and for all to the far more difficult verdict: *Ignorabimus*"—we shall never know.[24]

10.2 The Seven Enigmas

> When I was in Berlin I met du Bois-Reymond, and, wagging the end of my finger, I said to him, "What is that? What moves the finger?" He said he didn't know; that investigators have for twenty-five years been trying to find out. If anybody could tell him what wagged this finger, the problem of life would be solved.
>
> Thomas Edison, "What is Life?" 1891

Du Bois-Reymond's speech created a furor. Contemporaries likened it to an "unexpected explosion of a mine", coming as it did from "the center of the center of science".[25] Darwinists placed him in league with the "evil horde" of the Catholic Church.[26] Catholics regarded his argument as a challenge to the divinity of God.[27] Philosophers resented his trespass on grounds they considered their own.[28] And critics subjected him to abuse, sneering at his "incompetent ramblings", branding him as "the professor of national limitation", and charging him with offering the public

[20] Ibid., pp. 462–463. Most likely du Bois-Reymond is alluding to Heinrich von Kleist's version of *Amphitryon* (1807), in which the hapless Sosia is tricked and thrashed by his divine *Doppelgänger* Mercury.

[21] Ibid., p. 436.

[22] Ibid., pp. 463–464.

[23] Ibid., p. 462.

[24] Ibid., p. 464. "In medieval England a jury could bring in four alternative verdicts at a trial: Guilty, Not Guilty, *Ignoramus* (we do not know), *Ignorabimus* (we *shall* not know)." Humphrey 1982, p. 477.

[25] Anderton 1993, p. 215, quoting Kastan 1885; du Bois-Reymond 1885.

[26] Haeckel 1874, p. 131.

[27] Köstlin 1874, p. 18; Pesch 1875, p. 495.

[28] Hartmann 1876, p. 433.

"bisexual religious retrogressions, mysticisms, and multidimensional inanities".[29] For a long time du Bois-Reymond ignored his attackers, but after 8 years he had enough and answered them in a discourse titled "The Seven Enigmas".[30] The setting was significant: du Bois-Reymond addressed the Prussian Academy of Sciences on his home ground in Berlin. As one of the four permanent secretaries of the venerable institution he knew that his remarks would carry all the weight of his office.

Du Bois-Reymond opened the speech with an attack on the ignorance of his detractors. He confessed that he had hesitated before addressing his colleagues at Leipzig: the limits of science were well known to anyone familiar with the history of philosophy, and he had felt almost ashamed to offer them "so stale a draft".[31] His doubts had been misplaced. Most of his philosophical critics had assumed him to be a Kantian, a mistake in judgment that was A result of academic specialization.[32] "Since Kant transformed the discipline", du Bois-Reymond explained, "philosophy has taken on so esoteric a character, has so forgotten the language of common sense and plain thought, has so evaded the questions that most deeply stir our youth, or treated them condescendingly as officious speculations, and finally, has so opposed the rise of science, that it is not surprising that even the recollection of its earlier achievements has been lost".[33] In addition to forgetting the history of their own subject, philosophers also ignored metaphysics and religion, leaving many scientists to conclude that the field was empty.[34]

This condescending attitude blinkered the thinking of naturalists unacquainted with du Bois-Reymond's arguments. "Fanatics who should have known better" (meaning the morphologist Ernst Haeckel) "denounced me as belonging to the Black Band and demonstrated once again how near radicalism is to despotism. More temperate heads" (meaning the botanist Carl von Nägeli) "betrayed the weakness of their dialectics" in confusing the view that he endorsed, that consciousness was linked to material processes, with the view he opposed, that consciousness could be explained on a mechanical basis.[35]

The theologian David Friedrich Strauss was more insightful. Apart from saving du Bois-Reymond the trouble of "having to dash the hopes of those who mistakenly saw me as a champion of dualism", Strauss identified three questions that appeared to be insoluble:

(A) How has the living arisen out the lifeless?
(B) How has the sensible arisen out of the insensible?
(C) How has the conscious arisen out of the unconscious?

---

[29] Dietzgen 1903, p. 178. Hoffmann 1874, p. 504; Dühring 1878, p. 519; cf. Engelhardt 1976; Engelhardt 1981.

[30] du Bois-Reymond 1912g, pp. 93–94n1. My translations are based on du Bois-Reymond 1882.

[31] Ibid., p. 65.

[32] Dove 1898, p. 432.

[33] 1912g, 66; cf. Tennant 2007.

[34] Ibid., pp. 66–67.

[35] Ibid., p. 67.

Du Bois-Reymond viewed the second question as crucial, whereas Strauss inclined to the first and the third.[36] Here Strauss missed the point. Astronomical knowledge could indeed reveal the origin of life, since only the "wholly childish" insisted on successive periods of creation. The Almighty, du Bois-Reymond joked, was not some kind of amateur artist in continual need of improvement. Moreover, du Bois-Reymond had never asserted that sensation could explain consciousness; rather, he had asserted that the incomprehensibility of all higher mental processes "followed from it by an *a fortiori* argument." This accounted for why the "gap in our understanding" appeared at the second stage in Strauss's series and not after.[37]

Strauss thought that only time would tell if du Bois-Reymond's *Ignorabimus* would be the last word on the subject.[38] Du Bois-Reymond conceded that it was not, since Haeckel had taken his ironic suggestion that atoms were sentient and had spun it into a theory of inheritance that proposed the transmission of unconscious memories by means of "vivified atom-complexes".[39] At the very least, Haeckel's updated doctrine of anamnesis was obscure—one of his friends had studied it six times and still could not make sense of it.[40] It also lacked motivation. What was the point of attaching souls to atoms if mechanics explained our minds? Haeckel's error reminded du Bois-Reymond of the critics of Newton who associated gravity with will. "Whoever arrives at such nonsense", he scoffed, "and instead of humbly withdrawing, nails his colors to the mast and works himself into a frenzy of strident bombast, has indeed met success where Newton could only concede defeat".[41]

Having addressed the objections to his previous speech at Leipzig, du Bois-Reymond moved on to his current subject, "the seven shortcomings" of science.[42] He deemed the first two, the essence of matter and the origin of motion, to be inscrutably mysterious.[43] Continued investigation might well discover the origin of life, despite Pasteur's experience to the contrary, leaving a fourth problem in the apparently teleological arrangement of nature. Since morphological laws were inconsistent with the mechanical view, du Bois-Reymond regarded natural selection as the best answer to this conundrum.[44] By contrast, the fifth difficulty of the origin of sensation was quite transcendent.

Here du Bois-Reymond paused his discussion to review Leibniz's treatment of the issue. According to the German philosopher, "we are constrained to confess that perception and whatever depends upon it are inexplicable on mechanical principles; that is, by reference to forms and movements." "Imagine a machine", Leibniz

[36] Ibid., pp. 68–69; Strauss 1873, pp. xvi–xxiii.

[37] Ibid., p. 70.

[38] Strauss 1873, p. xxi.

[39] du Bois-Reymond 1912g, pp. 71–72; Haeckel 1876, pp. 38–39.

[40] Otto Zacharias to Ernst Haeckel, Dessau, 9 June 1876, in Nöthlich, et al. 2006, pp. 217–218; Emil du Bois-Reymond to Hermann Helmholtz, 7 May 1881, in Kirsten (Ed.) 2006, p. 264.

[41] du Bois-Reymond 1912g, pp. 71–73.

[42] English in the original.

[43] Ibid., p. 75.

[44] Ibid., pp. 75–76.

continued, "which manufactured thoughts, feelings, and perceptions, and think of it as enlarged in all its proportions, so that we could go into it as one might a mill. Even then we would find nothing but parts jostling each other, and never anything by which perception could be explained."[45] Du Bois-Reymond had followed the same line of reasoning in his essay on "The Limits of Science"; however, time had changed his mind. After all, it made a great difference whether charcoal, sulfur, and saltpeter were combined in large lumps or in a fine powder. He therefore rejected Leibniz's analogy and affirmed that consciousness could not be explained "as the result of any arrangement or motion of atoms"—a position, he added, that no one had attempted to challenge.[46]

Instead, his denigrators had contented themselves with making contradictory assertions, like Haeckel's charge that he had not considered the evolution of the human mind. This may have been the case, du Bois-Reymond allowed, but then Haeckel had failed to realize that our species had not altered since the time of Homer, that the world would freeze long before the advent of any super-beings, and that however much our brains might develop they could never surpass the powers of the Laplacian demon. "If anyone has sinned against evolution", du Bois-Reymond averred, "it is the Prophet of Jena".[47]

Du Bois-Reymond named the origin of intelligent thought and language as his sixth difficulty. He recognized the vast chasm between an amoeba and a person, but he expected the gap to be bridged in stages.[48] The "theory of knowledge" required only "memory and the power of generalization" to fashion complex thoughts out of simple sensations, and as great as was the intellectual divide between species, it paled in comparison with the rift between mechanics and mind. "To use to Strauss's notation again", he wrote, "if problem *B* is solved, problem *C* does not seem transcendent".[49]

With this du Bois-Reymond arrived at the final and the most important of the difficulties faced by science. Whereas the other problems in his list had been the concern of only a few intellectuals, the question of whether our actions were free—"touching everyone, apparently accessible to everyone, implicated with the fundamental conditions of society, impinging on the deepest religious convictions"—had played a part of "immeasurable moment in the history of ideas and civilization, and the stages of the development of the human mind" were plainly reflected in the discussion of it.[50] The earliest of these, classical antiquity, saw no contradiction between choice and necessity. Rather, it was Christian theology that complicated the question. If God was omnipotent, we were not free.

---

[45] Ibid., pp. 77–78.

[46] Ibid., p. 78. "Mancher hat Dubois-Reymond dies Beweisführung zum Vorwurf gemacht, Mancher seine Gebietsbegrenzung des Naturerkennens zu eng genannt, widerlegt hat ihn Keiner." Anon., *Leipziger Zeitung*, 25 January 1874, p. 45.

[47] Ibid., p. 79.

[48] Ibid., pp. 78–80.

[49] Ibid., pp. 80–81.

[50] Ibid., p. 80.

But then how could He hold us responsible for His will?[51] Du Bois-Reymond heaped scorn on the Church's reply.

> The doctrine of original sin, the questions of redemption through merit or through the blood of the Savior, by faith or by works, and of the different kinds of grace, were complicated in a thousand ways with that dilemma itself, already fruitful in subtleties, and the cloisters of Christendom resounded from the fourth to the seventeenth century with disputations about determinism and indeterminism. There is perhaps no subject of human consideration about which so many rows of untouched folios lie moldering away. But the controversy was not always confined to books. The bitter accusations of heresy that the ruling sect hurled at dissenters, with all their attendant horrors, hung all the more on such abstruse controversies the less they had to do with reason and the honest pursuit of truth.[52]

Science imagined freedom in altogether different terms. Force was neither created nor destroyed, which meant that everything was determined, including the molecules of our brains. The universe made most sense as a machine.[53]

Du Bois-Reymond pointed out that Leibniz had been the first to conceive of the world in this way. Freedom did not concern the philosopher, since God had ordained all things, including our experiences. However, Leibniz refused to accept we could ever find ourselves in the predicament of Buridan's ass (which starved because it could not choose between two equally distant piles of food) on the grounds that angels would tip the balance one way or the other. He also justified atrocities with the excuse that God had been forced to permit their existence in this best of all possible worlds.[54] Du Bois-Reymond found such metaphysics peculiar, to say the least, and endorsed only the objective side of Leibniz's determinism. Here he joined a long line of fatalists who regarded free will as an illusion, noting that we felt free in our dreams, that our waking thoughts seemed to come and go of themselves, and that much of the activity underlying purposeful movement went on beneath the level of our awareness.[55] Were our conscious actions really that much more deliberate?[56]

None of this seemed especially worrisome as long as we considered matters of minor importance. The trouble came when we imagined our ethical choices to be determined. "Even the most decided monist", du Bois-Reymond granted, "could hardly adhere to the earnest purposes of practical life in the face of the idea that all of human existence is a *fable convenue* in which mechanical necessity awards to Caius the part of a traitor, and to Sempronius that of a judge; and therefore Caius is taken to execution, while Sempronius goes to his breakfast. We are not bothered that so many letters in every hundred thousand miscarry because they are not addressed", "but it shocks our moral feelings to think that, according to Quetelet, so many persons in every hundred thousand are to become thieves, murderers, and arsonists; for

[51] Ibid., p. 81.

[52] Ibid., pp. 81–82.

[53] Ibid., p. 82.

[54] Ibid., pp. 83–85.

[55] As studies of reflex action and the autonomic nervous system had shown.

[56] Ibid., p. 85.

it is disconcerting to have to think that we are not criminals only because others, instead of ourselves, have drawn the black lots that might have fallen to our share".[57]

Du Bois-Reymond pointed out that most people—scientists, historians, judges, poets, dreamers, even those who just "sleepwalk through life"—saw no alternative but to ignore this dilemma.[58] Metaphysicians throughout history had attempted to reconcile morality with the mechanical view of the world; had they succeeded in squaring this circle, their attempts would have ceased. There was little hope of arriving at any solution *sub specie aeternitatis:* "only unconquerable problems tend to be this immortal".[59]

In the last section of his speech du Bois-Reymond appraised recent attempts to explain the interaction of mind and matter in terms of the singular solutions of certain differential equations. Three Catholic mathematicians had postulated that the soul could affect the motion of atoms in unstable equilibrium much as small disturbances could trigger avalanches.[60] Du Bois-Reymond countered that argument with the observation that even slight perturbations required the mind to perform work, an act that violated the conservation of energy. He therefore dismissed the claims of his French colleagues as a recrudescence of superannuated metaphysics, the atoms of the brain playing the part of Buridan's ass and the "directing principle" of the soul, whatever that might be, playing the part of Leibniz's angel.[61]

The seventh difficulty vanished if we denied free will; otherwise, it remained transcendent.[62] Du Bois-Reymond deemed it a poor consolation to monism to see dualism caught in the same net, "tangled all the more helplessly the more it struggles with ethics."[63] At one time he had thought that individual freedom was only a question of mechanics.[64] But later—and he made no secret of it—he experienced a Damascene moment.[65] During the winter of 1861 he came to believe "that at least three transcendental problems precede the problem of free will, namely, the nature of matter and force, the origin of motion, and the origin of sensation".[66]

---

[57] Ibid., p. 86. Caius Sempronius Gracchus was a liberal Roman senator who wanted to extend the franchise to all Latin citizens. His political opponent, Lucius Optimus, used the death of his servant as a pretext to arrest and execute thousands of Caius's supporters. The allusion to Bismarck's Anti-Socialist Law of 1878 was obvious.

[58] Ibid., pp. 86–87.

[59] Ibid., p. 87. 14. Cf. Figaro's philosophy, *The Marriage of Figaro*, act V, sc. 19: "Par le sort de la naissance,/L'un est roi, l'autre est berger:/Le hasard fit leur distance;/L'esprit seul peut tout changer./De vingt rois que l'on encense,/Le trépas brise l'autel;/Et Voltaire est immortel."

[60] Nye 1976, 280–281; Hacking 1983, 464–465; Hacking 1998, 150–159.

[61] du Bois-Reymond 1912g, pp. 87–91; Nye 1976, 290n38.

[62] Barrow 1998, pp. 232–236.

[63] du Bois-Reymond 1912g, p. 92.

[64] du Bois-Reymond 1848–1884, vol. 1, pp. xxxv–xxxvi.

[65] du Bois-Reymond 1912g, p. 93.

[66] du Bois-Reymond 1864, Bl. 32v; du Bois-Reymond to Hermann Helmholtz, 25 March 1862, in Kirsten (Ed.) 1986, pp. 202–203.

The enumeration of seven separate enigmas was merely an effect of the scientific division of labor. They might just as easily have been consolidated into "the enigma of the universe." Leibniz thought that he had resolved this problem, but had he listened to du Bois-Reymond's deliberations he would surely have agreed with his judgment of "*Dubitemus*": let us leave the question open.[67]

10.3 Sources and Significance

> The Saturnian once more took up the little mites, and Micromegas addressed them again with great kindness, though he was a little disgusted in the bottom of his heart at seeing such infinitely insignificant atoms so puffed up with pride. He promised to give them a rare book of philosophy, written in minute characters, for their special use, telling all that can be known of the ultimate essence of things, and he actually gave them the volume ere his departure. It was carried to Paris and laid before the Academy of Sciences; but when the old secretary came to open it, the pages were blank.
>
> "Ah!" said he. "Just as I expected."
>
> Voltaire, *Micromegas*, 1753

"The Seven Enigmas" caused a groundswell of outrage even larger than "The Limits of Science." Du Bois-Reymond met with obloquy from all across the world: scientists recoiled at his doubt, ecclesiastics castigated his usurpation, socialists decried his insolence, journalists slighted his language, writers parodied his arguments, and philosophers dismissed his logic.[68] The last response was particularly telling. There was no need to worry about the mystery of consciousness, Ernst Mach informed the Austrian Academy of Sciences, since "the problem was not a problem."[69] In this light the advent of positivism, pragmatism, and idealism at the end of the century can be seen as an effort to sweep du Bois-Reymond's argument under the rug.

Still, it would be a mistake to interpret the intensity of feeling generated by "The Limits of Science" and "The Seven Enigmas" as evidence that du Bois-Reymond wrote for his critics. In truth he did not think much of philosophy. He believed that it had much more to learn from science than science had to learn from it: experiments, as he wrote to a friend, had taught him to attend to fundamental incomprehensibility of the world.[70] The task of science was to reduce events to equations; where that failed, it was to delineate the limits of knowledge. As far as du Bois-Reymond was concerned his essays had done so satisfactorily, and he had come to a caesura in his

---

[67] du Bois-Reymond 1912g, p. 93.

[68] Hall 1881, p. 236; Mehring 1899–1900; Hartenau 1898; Bölsche 1918, pp. 79–80; Bölsche 1897, pp. 41–42; Bourget 1891, pp. 21–22; Theta 1882; Spir 1883, pp. 1–10; Ehrenfels 1886, pp. 483–484.

[69] Mach 1895, p. 208; cf. Lübbe 1981, pp. 140–141; Anderton 1993, pp. 501–503; Reichenberger 2007, p. 83.

[70] du Bois-Reymond 1912c, p. 438; Emil du Bois-Reymond to Gerhard Berthold 14 August 1874, in Dannemann 1919–1920, pp. 270–271; Herneck 1960, pp. 245–247.

thinking, “like a mathematician who demonstrates the impossibility of solving a problem”.[71] Any additional metaphysical speculation was on a par with astrology and alchemy.[72]

Neither were du Bois-Reymond’s essays intended as a sop to the Church. Since Haeckel’s allegations of Ultramontane conspiracy, critics had construed du Bois-Reymond’s *Ignorabimus* as a gambit for independence, or even worse, a Walk to Canossa.[73] Evidence points away from this. If du Bois-Reymond cared about the Church’s opinion of science, he would not have remarked that his opponents included “Catholic Jesuits fighting with open visors, and easily recognizable Protestant Jesuits fighting with closed”, nor would he have characterized Bishop Weber’s position as “one of a supernatural dualism which throws itself into the arms of Christian doctrine to the point of the Trinitarian dogma”.[74] Du Bois-Reymond’s lectures appear conciliatory only in the context of the *Kulturkampf*; viewed against the background of his career, they exhibit his abiding commitment to “Pyrrhonism in a new guise”.[75] Expressing doubt does not equate to admitting surrender. As Friedrich Lange remarked in his *History of Materialism*,[76] the rhapsodies of theologians and philosophers only denoted their conceit:

> Force and matter are inexplicable, models of atoms are only a “substitute” for true knowledge; therefore materialism is rejected—rejected by one of our top scientists. Why, then, can’t speculation and theology saunter onto the abandoned field and teach with great authority what science doesn’t know? (That they have no idea either doesn’t come into question). The celebrated physiologist has declared consciousness—indeed, the simplest sensation—inaccessible to research: why, then, shouldn’t good old metaphysics and faculty psychology drag out their puppets and set them dancing on the vacant field? The dreaded bugbear is gone; the scientist has sworn not to interfere; so the subject is ours again! Everything will carry on as if science didn’t exist.

Du Bois-Reymond did not need to add anything to this rejoinder.

The origins of “The Limits of Science” remain unclear. Some commentators identify Kant’s antinomies, Goethe’s sayings, or Müller’s agnosticism as German sources, but to my mind, Friedrich Schlegel’s analysis of irony with its use of paradox, its study of rhetoric, and its equation of actor and spectator, seems just as likely.[77] Themes of intellectual limitation appear in English. In 1860 John Tyndall wrote that

> when we endeavor to pass from the phenomena of physics to those of thought, we meet a problem which transcends any conceivable expansion of the powers which we now possess. We may think over the subject again and again, but it eludes all intellectual presentation. The territory of physics is wide, but it has its limits for which we look with vacant gaze into

[71] Emil du Bois-Reymond to Eugen Dreher, 3 October 1889, in Dreher 1900, pp. 113–115.

[72] Emil du Bois-Reymond to Gerhard Berthold, 14 August 1874, in Dannemann 1919–1920, pp. 270–271.

[73] Franck, “Erkenntnislehre” (1930–1931), p. 128; du Bois-Reymond 1974, p. xxxiii; Anderton 1993; Vidoni 1991, pp. 137–156; Reichenberger 2007.

[74] du Bois-Reymond 1907, pp. 7, 11.

[75] du Bois-Reymond 1912g, p. 94.

[76] Lange 1875, 2(2), 157–158.

[77] Wahsner 2007; Goethe 1850, *1*, 272; du Bois-Reymond 1974, p. xviii; Albert 1993.

> the region beyond. Let us follow matter to its utmost bounds, let us claim it in all its forms—even in the muscles, blood, and brain of man himself it is ours to experiment with and to speculate upon. Casting the term "vital force" from our vocabulary, let us reduce, if we can, the visible phenomena of life to mechanical attractions and repulsions. Having thus exhausted physics, and reached its very rim, a mighty Mystery still looms beyond us.[78]

This sounds a lot like the introduction to du Bois-Reymond's 1848 treatise, *Investigations in Animal Electricity*.[79] Tyndall claimed to have gotten the idea in Normandy while sitting under an elm, but his story seems about as plausible as the one about Newton's apple.[80] Du Bois-Reymond spent a day with Tyndall in London on 9 May 1855, just a few weeks before Tyndall's French reverie, and they could have discussed the limits of science then or at any time during their previous encounters.[81] To complicate matters further, the title of du Bois-Reymond's 1872 address recalled both Charles Kingsley's lecture on "The Limits of Exact Science as Applied to History" (1860) and Henry Longueville Mansel's meditations on *The Limits of Religious Thought* (1859). The whole question of English influence is vexed.[82]

Moreover, an equally good case can be made for French sources. The theme of intellectual limitation had been explored by Diderot, who was one of du Bois-Reymond's favorite authors, and by Voltaire, a bust of whom he kept in his living room.[83] It also had been mentioned by Bayle, La Mettrie, Condorcet, d'Alembert, and Rousseau, all of whom du Bois-Reymond read and cited, not to speak of Pascal, whose *Pensées* drew attention to both the incomprehensibility of matter and the "ataraxia, doubt, and perpetual suspension of judgment" of the Pyrrhonists.[84] Finally, it should be remembered that du Bois-Reymond's literary model, Sainte-Beuve, discussed all these luminaries in his *Causeries de lundi* (1948).

Du Bois-Reymond left one clue to the provenance of his determinism. In a letter to his parents dated 26 July 1838, written during the summer he discovered Lucretius, du Bois-Reymond recounted a debate with classmates in which he claimed that chance could not exist in a universe governed by physical law, and that

[78] Tyndall 1897.

[79] du Bois-Reymond 1848–1884.

[80] Barton 1987, p. 129.

[81] Emil du Bois-Reymond to Jeannette du Bois-Reymond, London, 31 Brook Street, 12 May 1855, SBBPK, Dep. 5 K. 11 Nr. 5.

[82] Du Bois heard Kingsley speak at the Royal Institution, and Mansel inspired Huxley's agnosticism. Emil du Bois-Reymond to Jeannette du Bois-Reymond, London, 24 April 1866, SBBPK, Dep. 5 K. 11 Nr. 5; Lightman 1987.

[83] Emil du Bois-Reymond to his parents, 25 June 1851, SBBPK, Dep. 5 K. 11 M. 5 Bl. 11; Ellen du Bois-Reymond, "El Arenal", SBBPK, Dep. 5, K. 12, Nr. 299, 27; Diderot 2000, p. 82; Denis Diderot to Sophie Volland, 15 October 1759, Diderot 1875–1877, *18,* pp. 407–409; Wahsner 2007, pp. 50–53; Voltaire 1901, pp. 227–231; Voltaire 1877–1883. Du Bois-Reymond tellingly claimed that Voltaire possessed the skeptical "spirit of the modern scientist, who never hesitates to concede his ignorance and to acknowledge the limits of his understanding." 1912a, p. 332.

[84] Rousseau 1762, *3*, p. 68; Pascal 1850, pp. 64, 66–67; Pascal 1995, p. 28; Vyverberg 1958; Baker 1975, p. 368; Wahsner 2007, pp. 47–50.

anything that we perceived as random was merely a transference of our own feelings of freedom onto a world of necessity.[85] A decade later he developed that idea in the introduction to his *Investigations in Animal Electricity*, arguing that force and matter were nothing other than anthropomorphic projections of ignorance, figures of speech that hypostatized "the same dualism which presents itself in the notions of God and the world, of soul and body, the same want which once impelled men to people bush and fountain, rock, air, and sea with creatures of their imagination".[86] It was here that he first addressed the problem of scientific limits:

> If we ask what is left if neither force nor matter possesses reality, those who stand with me at this point answer as follows: *It is simply not granted to the human mind to get beyond a final contradiction in these things*. We therefore prefer, instead of turning in circles of fruitless speculation, or hewing the knot asunder with the sword of self-delusion, to hold to the intuition of things as they are, to content ourselves, to use the poet's phrase, with the "wonder of what is there." For we cannot bring ourselves, by the simple reason that a true explanation eludes us in one direction, to shut our eyes to the defects of another, solely because no third alternative seems possible; and we possess enough renunciation to accept the idea that ultimately the one goal appointed to science may be not to comprehend the nature of things, but to comprehend that it is incomprehensible.[87]

This passage contains all the elements of du Bois-Reymond's later argument. The only difference between it and his mature philosophy was that in 1848 he still thought he could reconcile choice and constraint. That changed in the winter of 1861, when he began to assert there was no room for caprice in "the world of Epicurus," and that either one could look on history as Voltaire did, as an absurd *fable convenue*, or one could accept the harsh logic of Calvin's election of grace, which preserved the idea of providence at the cost of condemning apostates to be burned.[88]

10.4 The Famous Old Bear

> It is true that around every man a fatal circle is traced beyond which he cannot pass; but within the wide verge of that circle he is powerful and free.
>
> Alexis de Tocqueville, *Democracy in America*, 1835

Du Bois-Reymond should not have been surprised by the reaction to his speeches. The theme of forbidden knowledge had been handled in myth from the Garden of Eden to the Flight of Icarus, in literature from Dante's *Inferno* to Shelley's *Frankenstein*,

---

[85] Dep. 5 K. 10 Nr. 3, SBBPK, Dep. 5 K. 10 Nr. 3. Cf. Voltaire 1901, p. 231: "In effect, it would be very singular that all nature, all the planets, should obey eternal laws, and that there should be a little animal 5 feet high, who, in contempt of these laws, could act as he pleased, solely according to his caprice. He would act by chance; and we know that chance is nothing. We have invented this word to express the known effect of all unknown causes."

[86] Du Bois-Reymond 1848–1884*, 1*, pp. xl–xli.

[87] Ibid., *1*, pp. xlii-xlii. For a critique, see Aliotta 1914, p. 376. Kirchhoff's debt to du Bois-Reymond has been acknowledged only in passing. Oldham 2008, pp. 244–276.

[88] SBBPK, Nachlaß du Bois-Reymond, K. 12 M. 8 Nr. 11 Bl. 32^r–32^v; Bl. 37^v–38^r; Bl. 38^r–39^r.

and in philosophy from the Skeptics to the agnostics.[89] Du Bois-Reymond brought it up himself at least three times after the introduction to his *Investigations*: in his 1868 lecture on "Voltaire as a Scientist", where he referred to the limits of his understanding, in his 1870 lecture on "Leibnizian Ideas in Modern Science", where he mentioned the impossibility of comprehending even simple sensations, and in his 1872 lecture on the "History of Science", where he pointed out the areas in which science had reached the frontiers of its territory.[90] This did not mean that anyone took notice. Du Bois-Reymond's words hit home only after the creation of the state of Germany in 1871. There is nothing quite as unsettling as success.

Du Bois-Reymond's perceptions have stood up well. He questioned the possibility of understanding consciousness at a time when physiologists assumed that it arose naturally from the nervous system, he recognized the strengths and weaknesses of atomic models before they won general acceptance, he characterized life as a dynamic equilibrium at every level of organization, and he understood that science was necessarily flawed.[91] Indeed, the sophistication of his views has generally escaped historians of ideas, most notably Ernst Cassirer, whose *Determinism and Indeterminism in Modern Physics* caricatured his arguments as a muddle of contradictions.[92] What Cassirer failed to grasp was that quantum events are not necessarily random, and whereas some mathematicians believe that calculation cannot substitute for intuition, others suspect that intuition rests on an irremediable inconsistency in our thinking. Modern science does not entail Platonism.[93]

Du Bois-Reymond understood his critics far better than they understood him. His insight that metaphysics would cease to exist if consciousness were mechanically comprehensible refuted the idealist assumption that the mind lies beyond understanding.[94] Neither did he see any point in the phenomenalist program of reducing knowledge to sensation, since the ability to intuit things was precisely what was forbidden in a world without qualities. Moreover, the nominalist emphasis on the growing diversity of science only replaced coherent knowledge with an incongruous mess.[95] In contrast, du Bois-Reymond viewed mechanics as a heuristic. As Stephen Gaukroger has noted about Descartes, it was not that he thought

[89] Shattuck 1996.

[90] du Bois-Reymond 1912a, p. 332; 1912b, p. 388; 1912c, p. 437.

[91] Florey 1996, p. 168; Domin 1963, p. 115.

[92] Cassirer 1956, pp. 3–10, 48–49, 62–65, 149–152; Cassirer 2003, pp. 159–162 Cassirer took quantum mechanics for confirmation of idealism and positivism, which it isn't; he assumed du Bois-Reymond to be a naïve realist and materialist, which he wasn't; and worst of all, he passed off du Bois-Reymond's language and arguments as his own.

[93] Cushing 1994; Beller 1999; Stöltzner 2003; Howard 2004; Thurs 2009, pp. 196–205. For the relation of mechanism and consciousness, see Penrose 1989; Penrose 1994. For critiques, see McCarthy 1990; McCullough 1995; Chalmers 1995; Grush and Churchland 1995; LaForte, Hayes, and Ford 19981998. My favorite objection goes like this: Penrose's argument for the nonalgorithmicity of thought is entirely formal. One could imagine the Laplacian demon making it.

[94] du Bois-Reymond 1912g, p. 73; cf. 1912b, pp. 383–384.

[95] du Bois-Reymond 1912e, pp. 530–531.

that the world was a machine, rather that it was best imagined as a machine.[96] This is a subtle distinction, just as subtle as the politics of du Bois-Reymond's observation that the arbitrariness of history would disappear in a determinist universe.[97]

The impact of du Bois-Reymond's essays was so great that it is only now coming into view. His arguments raised an outcry that shaped debates over foundations in mathematics, measurement in physics, activity in neuroscience, will in psychology, and mind in philosophy.[98] Moreover, variants of his ideas showed up throughout modern culture – for example in Jean-Paul Sartre's meditation on the impossibility of self-knowledge, in Frederick Jackson Turner's thesis of the closing of the American frontier and in Henry Adams's search for unity as he wandered "through the forests of ignorance" and "necessarily fell upon the famous old bear that scared children at play".[99] In literature, du Bois-Reymond inspired Theodor Fontane, Heinrich Mann, Arthur Schnitzler, Hermann Broch, Robert Musil, Pío Baroja, Miguel de Unamuno, and, Gustave Flaubert, whose *Bouvard et Pécuchet* can be read as a rumination on the limits of knowledge.[100] Even the planet in Stanislaw Lem's *Solaris* is an embodiment of the Laplacian mind.[101]

A good deal of the difficulty of interpreting du Bois-Reymond's essays is conceptual. Du Bois-Reymond's faith in mechanism and liberalism trusted in the rule of law. Still, it does not follow that a decline in one ideology heralded a decline in the other, or that mechanism and liberalism declined at all.[102] As many narratives of change as there were at the end of the nineteenth century—and the shrillness of du Bois-Reymond's critics makes such announcements of revolution hard to ignore—there were also narratives of continuity. Instrumental approaches to science did not begin with Wilhelm Wundt and Ernst Mach. Emil du Bois-Reymond was perfectly aware of the elision of description and explanation.[103] By the same token, there is as much evidence for the strength of liberal values at the *fin-de-siècle* as there is for weakness.[104] Science never went bankrupt.

Ernst Cassirer identified du Bois-Reymond as the inventor of the doctrine of determinism.[105] Ian Hacking proved this wrong: Charles Renouvier discussed the concept in the 1850s, and if we consider Laplace's religious training a century ear-

---

[96] du Bois-Reymond 1912h, pp. 170–172; Gaukroger 2000, pp. 383–400.

[97] SBBPK, Nachlaß du Bois-Reymond, K. 12 M. 8 Nr. 11 Bl. 32[v]. *Willkür* means "arbitrariness" but can also translate as "despotism" or "caprice."

[98] Planck 1933, p. 118; Holton 1988; McGinn 1991; Nagel 1998.

[99] Sartre 1938; Turner 1921; Adams 1996, p. 429.

[100] Fontane 2002; Mann 1892; Riedel 1996, 231; Broch 1988, p. 55; Baroja 1974; 1920, p. 33; Unamuno 1925, p. 156; Flaubert 1976, p. 286.

[101] Lem 2002.

[102] Frank 2002, pp. 45–48.

[103] Wundt 1866; Mach 1911; du Bois-Reymond 1848–1884, *1*, xxv-l; du Bois-Reymond 1912h, pp. 170–171; Seth 2007, pp. 25–51.

[104] Beller 2001; Lees 2002; Jenkins 2003; Coen, 2007.

[105] Cassirer 1956, p. 3.

lier, the demon was nothing more than a variant of the Arminian belief in divine omniscience.[106] What is more, the idea that everything follows from natural causes dates back to Lucretius, or at the very least, to 1417, when Poggio discovered a copy of *De rerum natura* at Fulda. Contemporary philosophers continue to wrangle over the "hard problem of consciousness" without realizing that their arguments differ little from du Bois-Reymond's.[107]

Du Bois-Reymond said that the only thing to be learned from history is that there is nothing to learn. His doubt seems axiomatic to any study of the past. People who think there is a point to things tend to view history as a record of error, whereas those who are conscious of their ignorance tend to take the dead more seriously. Such an attitude of humility generally arises in the course of research. But it is also possible to come to this opinion by thinking. As du Bois-Reymond explained to his students, history was just the cosmic formula solved for negative values of time—a solution, he might have added, beyond the ken of all but a perfect mind.[108] Not that this was any great concern. "If the Laplacian demon should find anything in the cosmic formula obscure", du Bois-Reymond wrote to Carl Ludwig, "he need only take the train and come to Berlin".[109]

References

Adams H (1996) The education of Henry Adams: an autobiography. Modern Library, New York

Albert G (1993) Understanding irony: three *essais* on Friedrich Schlegel. MLN 108(5):825–848

Aliotta A (1914) The Idealist Reaction Against Science (trans: McCaskill A). Macmillan & Co, London

Anon. Zur kosmologischen Geistesbewegung der Gegenwart. Literarische Culturstudien. 2. Forschung, Wahn, Erkenntniß. Wissenschaftliche Beilage der Leipziger Zeitung, No. 8, Sunday, 25 Jan 1874, pp 41–45

Anderton KM (1993) The limits of science: a social, political, and moral agenda for epistemology in nineteenth century Germany. PhD dissertation, Harvard University

Baker KM (1975) Condorcet: from natural philosophy to social mathematics. University of Chicago, Chicago

Baroja y Nessi P (1920) Youth and Egolatry (trans: Fassett JS Jr, Phillips FL). Knopf, New York

Baroja y Nessi P (1974) The Tree of Knowledge (trans: Bell AFG). Fertig, New York

Barrow JD (1998) Impossibility: the limits of science and the science of limits. Oxford University, Oxford/New York

Barton R (1987) John Tyndall, pantheist: a rereading of the Belfast address. Osiris, 2nd ser., vol 3. History of Science Society, Philadelphia, pp 111–134

Beller M (1999) Quantum dialogue: the making of a revolution. University of Chicago, Chicago

Beller S (ed) (2001) Rethinking Vienna 1900. Berghahn Books, New York

[106] Hacking 1998, pp. 150–153.

[107] McCarty 2005; Tennant 2007.

[108] SBBPK, Nachlaß du Bois-Reymond, K. 12 M. 8 Nr. 11 Bl. 27^{v}, 29^{v}. "A dynamic theory would begin by assuming that all history, terrestrial or cosmic, mechanical or intellectual, would be reducible to this formula if we knew the facts." Adams 1996, p. 489.

[109] 26 November 1874, Cranefield 1982, p. 113.

Bölsche W (1918) Die Grenzen des Naturerkennens. Zum 100. Geburtstag du Bois-Reymonds. Reclams Univers Mod Illus Wochenschr 35(5):79–80
Bölsche W (1897) Du Bois-Reymond. Das Mag Litt 66(2):36–44
Bourget P (1891) Le Disciple. Lemerre, Paris
Broch H (1988) The Unknown Quantity (trans: Muir W, Muir E, 1933). Northwestern University, Evanston
Cassirer E (1956) Determinism and Indeterminism in Modern Physics: Historical and Systematic Studies of the Problem of Causality (trans: Benfey OT). Yale University/Oxford University, New Haven/London
Cassirer E (2003) Substance and Function, and Einstein's Theory of Relativity (trans: Swabey WC, Swabey MC, 1923). Dover, New York
Chalmers DJ (1995) Minds, machines, and mathematics: a review of *Shadows of the mind* by Roger Penrose. Psyche 2(9), n.p.
Coen DR (2007) Vienna in the age of uncertainty: science, liberalism, and private life. University of Chicago, Chicago
Cranefield P (ed) (1982) Two great scientists of the nineteenth century: correspondence of Emil du Bois-Reymond and Carl Ludwig, collected by Estelle du Bois-Reymond, foreword, notes, and indexes by Paul Diepgen (trans: Sabine Lichtner-Ayèd). Johns Hopkins University, Baltimore/London
Cushing JT (1994) Quantum mechanics: historical contingency and the Copenhagen hegemony. University of Chicago, Chicago
d'Alembert J (1893) Discours préliminaire de l'Encyclopédie (1751). Delagrave, Paris
Dannemann F (1919–1920) Aus Emil du Bois-Reymonds Briefwechsel über die Geschichte der Naturwissenschaften. Mitt Gesch Med Naturwiss 18:267–274; 19: 1–8
Diderot D (1875–1877) Oeuvres complètes de Diderot. Assésat J, Tourneux M (eds). Garnier Frères, Paris
Diderot D (2000) Lettre sur les aveugles (1749). Hobson M, Harvey S (eds). Flammarion, Paris
Dietzgen J (1903) Unsere Professoren auf der Grenzen der Erkenntnis (1878). In Kleinere philosophische Schriften. Eine Auswahl. Dietz, Stuttgart, pp 162–178
Domin G (1963) Einige philosophiehistorische Fragen zu den theoretischen Auseinandersetzungen Emil du Bois-Reymonds. In: Harig G, Mette A (eds) Naturwissenschaft, Tradition, Fortschritt. Beiheft zu NTM: Zeitschrift für Geschichte der Naturwissenschaft, Technik und Medizin. VEB Deutscher Verlag der Wissenschaften, Berlin, pp 112–118
Dove AW (1898) Der neue Glaube nach David Strauß (1872). In: Ausgewählte Schriftchen, vornehmlich historischen Inhalts. Duncker & Humblot, Leipzig, pp 426–452
Dreher E (1900) Die Grundlagen der exakten Naturwissenschaft im Lichte der Kritik. Apollo, Dresden
du Bois-Reymond E (1848–1884) Untersuchungen über thierische Elektricität. Reimer, Berlin
du Bois-Reymond E (1864) Findings of contemporary science. Staatsbibliothek Preußischer Kulturbesitz zu Berlin, Haus 2, Handschriftenabteilung, *Nachlaß* du Bois-Reymond, K. 12 M. 8 Nr. 11
du Bois-Reymond E (1874) The Limits of the Knowledge of Nature (trans: Fitzgerald J). Pop Sci Mon 5(2):17–32
du Bois-Reymond E (1882) The seven world-problems. Pop Sci Mon 20(28):433–447
du Bois-Reymond, E (1885) Zum Rektoratwechsel der Berliner Universitat am 15. Oktober. Berliner Fremdenblatt, Nr. 242, 15 October, 4
du Bois-Reymond E (1907) Über die Grenzens des Naturerkennens. Die sieben Welträthsel. Zwei Reden. Veit, Leipzig
du Bois-Reymond E (1912a) Voltaire als Naturforscher. In der Friedrichs-Sitzung der Akademie der Wissenschaften am 30. Januar 1868 gehaltene Rede. In: Reden, 1. Veit, Leipzig, pp 319–348
du Bois-Reymond E (1912b) Leibnizische Gedanken in der neueren Naturwissenschaft. In der Leibniz-Sitzung der Akademie der Wissenschaften am 7. Juli 1870 gehaltene Rede. In: Reden, vol 1. Veit, Leipzig, pp 370–392

du Bois-Reymond E (1912c) Über Geschichte der Wissenschaft. In der Leibniz-Sitzung der Akademie der Wissenschaften am 4. Juli 1872 gehaltene Rede. In: Reden, vol 1. Veit, Leipzig, pp 431–440
du Bois-Reymond E (1912d) Über die Grenzen des Naturerkennens. In der zweiten allgemeinen Sitzung der 45. Versammlung Deutscher Naturforscher und Ärzte zu Leipzig am 14. August 1872 gehaltener Vortrag. In: Reden, vol 1. Veit, Leipzig, pp 441–473
du Bois-Reymond E (1912e) La Mettrie. In der Friedrichs-Sitzung der Akademie der Wissenschaften am 28. Januar 1875 gehaltene Rede. In: Reden, vol 1. Veit, Leipzig, pp 509–539
du Bois-Reymond E (1912f) 'Aus den Llanos'. Anzeige und Nekrologe (1879). In: Reden, vol 2. Veit, Leipzig, pp 46–64
du Bois-Reymond E (1912g) Die sieben Welträtsel. In der Leibniz-Sitzung der Akademie der Wissenschaften am 8. Juli 1880 gehaltene Rede. In: Reden, vol 2. Veit, Leipzig, pp 65–98
du Bois-Reymond E (1912h) Goethe und kein Ende. In der Aula der Berliner Universität am 15. Oktober 1882 gehaltene Rektoratsrede. In: Reden, vol 2. Veit, Leipzig, pp 157–183
du Bois-Reymond E (1974) Vorträge über Philosophie und Gesellschaft. Wollgast S (ed). Meiner, Hamburg
Dühring E (1878) Kritische Geschichte der Philosophie von ihren Anfängen bis zur Gegenwart, 3rd edn. Fues, Leipzig
von Ehrenfels C (1886) Metaphysische Ausführungen im Anschlüsse an Emil du Bois-Reymond. Sitzung Philos-Hist Klasse Kais Akad Wiss Wien 112:429–503
von Engelhardt D (1976) Du Bois-Reymond 'Über die Grenzen des Naturerkennens'—eine naturwissenschaftliche Kontroverse im ausgehenden 19. Jahrhundert. Orvostör Közl (Commun Hist Artis Med) 80:9–25
von Engelhardt D (1981) Du Bois-Reymond im Urteil der zeitgenössischen Philosophie. In: Mann G (ed) Naturwissen und Erkenntnis im 19. Jahrhundert: Emil du Bois-Reymond. Gerstenberg, Hildesheim, pp 187–205
Finkelstein G (2013) Emil du Bois-Reymond: Neuroscience, Self, and Society in Nineteenth-Century Germany. MIT, Cambridge/London
Flaubert G (1976) Bouvard and Pécuchet (trans: Krailsheimer AJ). Penguin, London/New York
Florey E (1996) Das 5. Welträtsel—Ignorabimus? Über die Unmöglichkeit, bewußtes Empfinden physiologisch zu erklären. Vortrag auf der Festveranstaltung anläßlich des 100. Todestages von Emil du Bois-Reymond (1818–1896) am 14. Dezember 1996 in Berlin. In: Festschrift zum 100. Todestag von Emil du Bois-Reymond (1818–1896). Akademie, Berlin, pp 161–174
Fontane T (2002) Effi Briest. Reclam, Stuttgart
Franck P (1930–1931) Was bedeuten die gegenwärtigen physikalischen Theorien für die allgemeine Erkenntnislehre? Erkenntnis 1:26–157
Frank P (2002) Einstein: His Life and Times (trans: Rosen G). Da Capo, New York
Gaukroger S (2000) The resources of a mechanist physiology and the problem of goal-directed Processes. In: Gaukroger S, Schuster J, Sutton J (eds) Descartes' natural philosophy. Routledge, London, pp 383–400
Goethe J (1850) Conversations of Goethe with Eckermann and Soret (trans: John Oxenford). Smith, Elder, and Co, London
Grush R, Churchland PS (1995) Gaps in Penrose's toilings. J Conscious Stud 2(1):10–29
Hacking I (1983) Nineteenth century cracks in the concept of determinism. J Hist Ideas 44(3):455–475
Hacking I (1998) The taming of chance. Cambridge University, Cambridge/New York
Haeckel E (1874) Anthropogenie, oder, Entwickelungsgeschichte des Menschen. Gemeinverständliche wissenschaftlich Vorträge über die Grundzüge der menschlichen Keimes- und Stammes-Geschichte. Engelmann, Leipzig
Haeckel E (1876) Die Perigenesis der Plastidule, oder die Wellenzeugung der Lebenstheilchen. Ein Versuch zur mechanischen Erklärung der elementaren Entwickelungs-Vorgänge. Reimer, Berlin
Hall GS (1881) The muscular perception of space. In: Aspects of German culture. Osgood, Boston, pp 208–236

Hartenau W [Walter Rathenau] (1898) Ignorabimus. Die Zuk 25:524–536
von Hartmann E (1876) Naturforschung und Philosophie (1867). In: Gesammelte Studien und Aufsätze gemeinverständlichen Inhalts. Duncker, Berlin, pp 421–444
Herneck F (1960) Emil du Bois-Reymond und die Grenzen der mechanistischen Naturauffassung. In: Forschen und Wirken. Festschrift zur 150-Jahr-Feier der Humboldt-Universität zu Berlin, 1810–1960, vol 1. VEB Deutscher Verlag der Wissenschaften, Berlin, pp 229–251
Hoffmann F (1874) Über die Grenzen des Naturerkennens. Zwei Recensionen. Psych Stud 1(11):502–508
Holton G (1988) The roots of complimentarity. In: The thematic origins of scientific thought: Kepler to Einstein. Harvard University, Cambridge, pp 99–146
Howard D (2004) Who invented the Copenhagen interpretation? A study in mythology. Philos Sci 71(5):669–682
Humphrey N (1982) Consciousness: a just-so story. New Sci 95(1319):474–477
Jenkins J (2003) Provincial modernity: local culture and liberal politics in fin-de-siècle Hamburg. Cornell University, Ithaca
Kastan I (1885) Studien-Erinnerungen. II. Emil Du Bois-Reymond. Nat-Ztg 38(528), 20 Sept, Sonntags-Beilage
Kingsley C (1860) The limits of exact science as applied to history. An inaugural lecture delivered before the University of Cambridge. Macmillan, Cambridge/London
Kirsten C (ed) (1986) Dokumente einer Freundschaft. Briefwechsel zwischen Hermann von Helmholtz und Emil du Bois-Reymond, 1846–1894. Akademie, Berlin
Köstlin O (1874) Ueber die Grenzen der Naturwissenschaft, 2nd edn. Fues, Tübingen
LaForte G, Hayes PJ, Ford KM (1998) Why Gödel's theorem cannot refute computationalism. Artif Intell 104:265–286
Lange FA (1873–1875) Geschichte des Materialismus und Kritik seiner Bedeutung in der Gegenwart, 2nd edn. Baedeker, Iserlohn
Laplace PS (1814) Essai philosophique sur les probabilités, 3rd edn. Courcier, Paris
Lees A (2002) Cities, sin, and social reform in imperial Germany. University of Michigan, Ann Arbor
Lem S (2002) Solaris (trans: Kilmartin J, Cox S). Harcourt Brace, San Diego
Lightman B (1987) The origins of agnosticism: Victorian unbelief and the limits of knowledge. The Johns Hopkins University, Baltimore/London
Lübbe H (1981) Wissenschaft und Weltanschauung. Ideenpolitische Fronten in Streit um Emil du Bois-Reymond. In: Mann G (ed) Naturwissen und Erkenntnis im 19. Jahrhundert: Emil du Bois-Reymond. Gerstenberg, Hildesheim, pp 129–148
Mann H (1892) Neue Romantik. Die Ggw 42(29):40–42
McCarthy J (1990) Review of *The emperor's new mind*. Bull Am Math Soc 23(2):606–616
McCarty DC (2005) Problems and riddles: Hilbert and the Du Bois-Reymonds. Synthese 147(1):63–79
McCullough D (1995) Can humans escape Gödel? A review of *Shadows of the mind* by Roger Penrose. Psyche 2(4), n.p.
McGinn C (1991) The problem of consciousness: essays toward a resolution. Blackwell, Oxford/Cambridge
Mach E (1895) The economical nature of physical inquiry (1882). In: Popular Scientific Lectures (trans: McCormack TJ. Open Court, Chicago, pp 186–213
Mach E (1911) History and root of the principle of the conservation of energy. Open Court, Chicago
Mansel HL (1859) The limits of religious thought examined in eight lectures delivered before the University of Oxford, in the year MDCCCLVIII., on the Bampton Foundation. Gould and Lincoln, Boston
Mehring F (1899–1900) Review of Ernst Haeckel. Die Welträtsel Die Neue Zeit 18(1):417–421
Mill JS (1859) Bain's *Psychology*. Edinb Rev 110(224):287–321
Nagel T (1998) Conceiving the impossible and the mind-body problem. Philosophy 73(285): 337–352

Nöthlich R, Wetzel N, Hoßfeld U, Olsson L (2006) 'Ich acquirirte das Schwein sofort, ließ nach dem Niederstechen die Pfoten abhacken u. schickte dieselben an Darwin'—Der Briefwechsel von Otto Zacharias mit Ernst Haeckel (1874–1898). Ann Hist Philos Biol 11:177–248

Nye MJ (1976) The moral freedom of man and the determinism of nature: the Catholic synthesis of science and history in the *Revue des questions scientifiques*. Br J Hist Sci 9(3):274–292

Oldham KTS (2008) The doctrine of description: Gustav Kirchhoff, classical physics, and the "purpose of all science" in 19th-century Germany. PhD dissertation, University of California, Berkeley

Pascal B (1850) The Thoughts on Religion, and Evidences of Christianity (trans: Pearce G). Longman, Brown, Green, and Longmans, London

Pascal B (1995) Pensées and Other Writings (trans: Levi H). Oxford University, Oxford/New York

Penrose R (1989) The emperor's new mind: concerning computers, minds, and the laws of physics. Oxford University, Oxford

Penrose R (1994) Shadows of the mind: a search for the missing science of consciousness. Oxford University, Oxford

Pesch T (1875) Philosophische Bestrebungen im deutschen Kulturlager. Stimmen Maria-Laach 8:481–495

Planck M (1933) Kausalgesetz und Willensfreiheit. Öffentlicher Vortrag, gehalten in der Preußischen Akademie der Wissenschaften am 17. Februar 1923. In: Wege zur physikalischen Erkenntnis. Reden und Vorträge. Hirzel, Leipzig, pp 87–127

Reichenberger A (2007) Emil Du Bois-Reymonds Ignorabimus-Rede: ein diplomatischer Schachzug im Streit um Forschungsfreiheit, Verantwortung und Legitimation der Wissenschaft. In: Bayertz K, Gerhard M, Jaeschke W (eds) Weltanschauung, Philosophie und Naturwissenschaft im 19. Jahrhundert. Band 3: Der Ignorabimus-Streit. Meiner, Hamburg, pp 63–87

Riedel W (1996) "Homo Natura". Literarische Anthropologie um 1900. de Gruyter, Berlin/New York

Rousseau J-J (1762) Émile, ou, de l'éducation. Jean Néaulme, Amsterdam

Sainte-Beuve C-A (1948) Causeries de lundi. Librairie Garnier Frères, Paris

Sartre J-P (1938) La Nausée. Gallimard, Paris

Seth S (2007) Crisis and the construction of modern theoretical physics. Br J Hist Sci 40(1):25–51

Shattuck R (1996) Forbidden knowledge: from Prometheus to pornography. St. Martin's, New York

Spir A (1883) Studien. Findel, Leipzig

Stöltzner M (2003) Causality, realism, and the two strands of Boltzmann's legacy (1896–1936). PhD dissertation, Bielefeld University

Strauss DF (1873) The Old Faith and the New: A Confession (trans: Mathilde Blind). Henry Holt and Company, New York

Tennant N (2007) Mind, mathematics and the *Ignorabimusstreit*. Br J Hist Philos 15(4):745–753

Theta (1882) Das Weltproblem. Neue Freie Presse. Abendblatt, Nr. 6305, Vienna, Thursday, 16 March. Staatsbibliothek Preußischer Kulturbesitz zu Berlin, Haus 2, Handschriftenabteilung, *Nachlaß* du Bois-Reymond, K. 5. M. 6 Nr. 24

Thurs DP (2009) Myth 22: that quantum physics demonstrated the doctrine of free will. In: Numbers RL (ed) Galileo goes to jail and other myths about science and religion. Harvard University, Cambridge/London, pp 196–205

Turner FJ (1921) The frontier in American history. Holt, New York

Tyndall J (1897) Physics and metaphysics. Saturday Review, 4 Aug 1860. In: Fragments of science: a series of detached essays, addresses, and reviews, vol 2. Appleton, New York, pp 393–394

de Unamuno M (1925) My religion. In: Essays and Soliloquies (trans: Flitch JEC). Knopf, New York, pp 154–162

Vidoni F (1991) Ignorabimus! Emil du Bois-Reymond und die Debatte über die Grenzen wissenschaftlicher Erkenntnis im 19. Jahrhundert. Peter Lang, Frankfurt/Bern/New York/Paris

Voltaire (1877–1883) Les colimaçons du Révérend Père L'Escarbotier, par la grâce de Dieu capucin indigne prédicateur ordinaire et cuisinier du grand couvent de la ville de Clermont en Auvergne (1768). In: Oeuvres complètes de Voltaire. Nouvelle edition, vol 27. Garnier Frères, Paris, pp 213–226

Voltaire (1901) The ignorant philosopher (1766). In: The Works of Voltaire: A Contemporary Version (trans. Fleming WF), vol 35. DuMont, Paris/London/New York/Chicago, pp 219–288
Vyverberg H (1958) Historical pessimism in the French enlightenment. Harvard University, Cambridge
Wahsner R (2007) Debatten über die Grenzen des Naturerkennens vor dem Ignorabimus-Streit. In: Bayertz K, Gerhard M, Jaeschke W (eds) Weltanschauung, Philosophie und Naturwissenschaft im 19. Jahrhundert. Band 3: Der Ignorabimus-Streit. Meiner, Hamburg, pp 36–62
Wundt W (1866) Die physikalischen Axiome und ihre Beziehung zum Causalprincip. Ein Capitel aus einer Philosophie der Naturwissenschaften. Enke, Erlangen

Chapter 11
William James and the "Theatre" of Consciousness

Stephanie L. Hawkins

> *To clear up the* immediate *relation of minds and brains is the big problem.*
>
> –William James

Such was the "hard problem" of consciousness as James described it in an entry in a notebook from 1905, penned just 5 years before his death while he tried in vain to complete his work-in-progress on a coherent and systematic philosophy.[1] While writing *The Principles of Psychology* 17 years earlier, a frustrated James railed against the hard problem in a letter to a colleague: "experience is cognitive, but who, what, or where the vehicle of cognition is, transcends my powers!".[2] In fact, James's career-long interest in such exceptional mental states as religious experience, hysteria, and mediumistic trance can arguably be understood as an effort to develop an evidential body of subjective data to form the basis for the more systematic mind philosophy he was working toward before his death. As I have written elsewhere, the radically empiricist, pluralistic philosophy James devised in his later years fused the nomenclature of evolutionary biology with that of the "new" physics, an invisible realm of "supersensual" reality, disclosed by magnetic fields and invisible penetrating rays, whose erratic behavior resembled the unpredictable course of consciousness itself.[3] As James would famously say of consciousness, "motion there obeys no Newton's laws."[4] In so doing, James anticipates contemporary accounts in the mind sciences that draw upon both quantum physics and phenomenology in order to

[1] "Notebook 19." bMSAm1092.9 (4513). William James Papers (MSAm1092.9–1092.12). Houghton Library, Harvard University.

[2] Skrupskelis 1995, p. 75.

[3] Hawkins 2011.

[4] James 1922, p. 34.

S.L. Hawkins (✉)
Department of English, University of North Texas, Denton, TX, USA
e-mail: shawkins@unt.edu

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_11,

address the unique challenges of identifying the uncertain physiological origins of "first-person" conscious experience, what many in the mind sciences have acknowledged to be among the "hardest" of the hard problems David Chalmers has delineated.[5]

In the case of the contemporary mind sciences, resolving the "explanatory gap" between one's internal self awareness or "consciousness" and brain physiology involves not only a rethinking of the fundamental components of reality and of the mind's engagements with the natural world, but also invites a revitalized understanding of how the very architecture of the brain gives rise to conscious thought.[6] Both modes of inquiry suggestively point to the need for more concerted interdisciplinary collaborations between evolutionary biologists, quantum physicists, and phenomenologists for a potential resolution to the hard problem. As a painter by vocation, natural scientist by training, and a philosopher by inclination, James's early education predisposed him to transgress conventional disciplinary boundaries in his pursuit of the stubborn facts of human psychology.[7]

In addition to highlighting James's pioneering contributions to the history of the "hard problem" in neuroscience, this essay aims to deepen our understanding of James's interdisciplinary fusion of Darwin, physics, and phenomenology in his effort to address the relation of our felt sense of possessing an autonomous, conscious "self" to a broader environment and social community. In order to address this central concern of James's, I first touch on his earliest writings describing consciousness as a selective agency that works in consort with natural selection to contribute to human evolutionary advancement. Second, I will show how James grafted Darwinian evolutionary thought onto the new physics in ways that anticipate contemporary neuroscientific attempts to resolve the hard problem by identifying the structures in the brain responsible for our phenomenological experience of qualia – those uniquely sensory, temporal, and spatial attributes of the world that we have come to associate with our sense of possessing a finite, yet agential "self." Finally, I will address the broader sociological aspects of the hard problem in James's later writings. While James's late philosophical writings may at first appear to have nothing to do with the "hard problem" *per se*, I maintain that if we consider the hard problem in terms of its implications for scientific inquiry and society itself, we will see James addressing these broader consequences. James was a pioneer in the realm of neuroscience; as well as founding the first experimental psychology laboratory in the U.S., he remained devoted to developing a theoretical framework for understanding consciousness as both an emergent property of complex, dynamical brain functions and as a transformative social force. Through the

[5] Chalmers 2010. Throughout this essay I will use the term, the "mind sciences" to refer collectively to philosophers of mind, neuroscientists, and psychologists and physicists invested in exploring the relationship of consciousness to that of brain physiology.

[6] Thompson 2007; Revonsuo 2009; Fingelkurts et al. 2010.

[7] Cotkin 1990; Bordogna 2008.

lens of James's later writings, we can understand the hard problem as encompassing the biosocial: the experience of individual and collective minds in the process of shaping complex networks of social interaction.

11.1 Volition: The Selective Agency of Consciousness

The "hard problem" of reconciling brain physiology to the subjective sensation of having a conscious self-awareness, or "self," was a problem with which James grappled almost from the beginning of his psychological and philosophical career. Based on his avid reading of physics, James correlated consciousness with the concepts of "energy" and "force," analogies that remained potent throughout his writings on consciousness. James lent both of these concepts a psychological dimension by fusing philosophical idealism with British and Germanic strains of scientific materialism. Will, habit, and attention form the Jamesean triumvirate of psychological inquiry concerning subjectivity and volition, dating from his earliest days as a student.[8] Between 1864, the year James enrolled at Harvard's medical school, and 1875, the year he launched Harvard's and the nation's first experimental psychology laboratory, he had read Laplace in mathematics; Newton, Maxwell, and Planck in physics; Hughlings-Jackson in neurology; Spinoza, Leibnitz, Descartes, and Schopenhauer in philosophy; and Galton, Spencer, and Wundt in psychology.[9] He was well versed in neurology and was deeply influenced by German laboratory science, for he had studied physiology and experimental neurology in Berlin, and experimental psychology at Heidelberg under Wundt and Helmholtz. In 1875, the same year James opened his experimental psychology laboratory at Harvard, he also gave a series of ten lectures at Johns Hopkins on "The Brain and the Mind." In the midst of all his heady reading in physics, James also read Charles Darwin on the origin of species and Jonathan Edwards on original sin. Natural science, physics, and Christian theology supplied the intellectual and philosophical ballast for James's understanding of consciousness as a component of the personal subjectivity known as the "self," and the cultivation of the will, which formed the basis for individual beliefs and subsequent actions. The study of what motivates individual group choices and actions, then, forms the core of James's person-centered investigation of consciousness.

The evolutionary trend of James's thought pervades his earliest writings on volition. James's discussion of "will" had Lamarckian overtones in his emphasis on the idea that consciousness could be cultivated and enlarged, and, furthermore, that it directs the brain, shapes its physiology, and, by assigning it certain tasks, thereby strengthens it.[10] Two of his earliest essays, "Are We Automata?" (1879)

[8] Feinstein 1999; Leary 2002.

[9] Taylor 1996, p. 73.

[10] Crippen 2010.

and "What the Will Effects" (1888), establish the twin concepts "will" and "force," as consistently denotative of the selective principle of consciousness itself, that inform James's writings. Ultimately, the Jamesean concept of consciousness as a selective agency was staked upon his philosophical ethics and on the role of choice in the gradual building up of character and of an individual "self."

In the first of these essays, "Are We Automata?" (1879/1983) James delineates the selective agency of consciousness in relation to a nascent phenomenological approach to resolving the hard problem. From James's perspective, consciousness, by responding to an individual's interests, "decides" what objects will occupy the center and what objects will drift toward the margins of perception. While T. H. Huxley's "Conscious-Automaton-Theory" posited consciousness as a "supernumary" epiphenomenon incidental to the brain, James, on the other hand, asserted the utility of consciousness on the basis of Darwinian natural selection. By exercising its selective agency in attending to what is of vital interest to individual survival, consciousness, James maintained, lends evolutionary processes greater efficiency. To underscore the point, James deployed metaphors emphasizing the "aesthetic" condition of consciousness's selective role. If the brain is a "machine," then consciousness is a "sculptor" or "painter."[11] James thus assigns consciousness's invisibly selective movements – at least from the point of view of Huxley's positivism – a weight and substance, in so far as conscious selection, as an aspect of the brain, has the potential to move and shape itself and its environment. In the passage from which this essay takes its title, James writes,

> the mind is at every stage a theatre of simultaneous possibilities. Consciousness consists in the comparison of these with each other, the selection of some, and the suppression of the rest by the reinforcing and inhibiting agency of Attention. The highest and most elaborated mental products are filtered from the data chosen by the faculty next beneath out of the mass offered by the faculty below that, which mass in turn was sifted from a still larger amount of yet simpler material, and so on. The highest distillate thus *represents* in the last analysis nothing but sensational elements. But this is far from meaning that it implies nothing but passive faculty of sensation. As well might one say that the sculptor is passive, because the statue stood from eternity within the stone. So it did, but with a million different ones beside it. The world as Goethe feels and knows it all lay embedded in the primordial chaos of sensations, and into these elements we may analyze back every thought of the poet. We may even, by our reasonings, unwind things back to that black and jointless continuity of space and moving clouds of swarming atoms which science calls the only real world.[12]

This early passage, which James would later incorporate into his chapter on "The Stream of Consciousness" in his 1890 *Principles of Psychology*, already contains both the Darwinian and psychophysical germs of James's pluralistic philosophy. The selective agency of consciousness shapes the "primordial chaos of sensations" into a coherent internal world. Furthermore, this "primordial" substance of which the entire universe is composed contains all the latent possibilities for multitudinous consciousnesses, from Goethe and "cuttlefish and crab" to the present age.

[11] James 1879/1983, pp. 51–52.

[12] James 1981, p. 51.

Crucially, moreover, this passage indicates that James had come to view volitional "will" and habitual, or repeated, acts of perceptual "attention" as coterminous with consciousness itself.

In this same essay, James ascribes to consciousness a hierarchical evolutionary structure, whereby consciousness "ascends" from lower and simpler forms to successively higher and more complex states. This structure invokes both the Spencerian notions of progress from homogeneous to increasingly complex, heterogeneous forms, and the Darwinian principle of natural selection. What is at stake for James is the building up of character through the exercise of habit, attention, and will in the process of self-making, with consciousness acting as a selective agency that harnesses the energies and processes of the physical brain. According to James's hypothesis, consciousness, through the agency of selection, gives force and focus to processes that would otherwise be random and scattershot. Over successive generations, natural selection may have led the human organism to its auspicious capacity for higher reasoning, James maintains, but human reasoning itself can exercise an even greater selective process by willfully directing the mind's attention, thereby training the mind to adapt to different situations and experiences, which, in turn, shape or selectively cultivate the neuronal pathways of the brain.[13]

James's "What the Will Effects" (1888/1983), further elaborates on his Darwinian description of volition as a "feeling" of "effort" that is engaged when one experiences a "struggle" to maintain attention or focus. From here, James develops his theory that consciousness is a "stream" wherein we perceive events flowing one into the other, creating our sense of wholeness. We arrest the flow of our conscious stream in order to introspectively analyze its contents. Herein lies the origin of what has been termed James's "field theory" of consciousness, taken up by recent mind scientists and philosophers as a useful framework for establishing a culture-independent structure for human consciousness.[14]

From a history of neuroscience perspective, what is most significant in the essay "What the Will Effects" is the role of volition articulated by James's concept of will as a "forcible holding fast to an incongenial idea," and the ensuing violent Darwinian struggle that takes place psychologically. James correlates consciousness with selection, and, specifically, with attention by arguing that volition and attention amount to the same things. In so doing, James thus lends ideas themselves the "force" to move individuals, which makes James's psychology significant in the context of the Progressive Era and its preoccupation with social reform.[15] Anticipating James's Pragmatism, "What the Will Effects" represents consciousness's selective agency as one that "wills" based on moral choice. This is a very different position from Huxley's biological model of reflex actions, yet James

[13] James, of course, anticipates Gerald Edelman's theory of "neural Darwinism." However, Edelman parts ways with James by insisting that consciousness itself has no impact on neuronal firing. See Edelman 2004, p. 84. See also, Schwartz and Begley *The Mind and the Brain* (2003), which cites James on the adaptive "plasticity" of the brain in his chapter on "Habit" in *Principles*.

[14] See Revonsuo 2009, Fingelkurts et al. 2010, and Barnard 1997.

[15] James 1879/1983, pp. 229–230.

retains the Darwinian notion of evolutionary struggle in his description of a struggle between competing internal psychological needs or goals and external environmental and societal forces. As James has it, our individual wills can either be supported or unsupported by the environment; furthermore, our will can either be supported or unsupported by our social context as much as it can be either affirmed or undermined by both the social and the natural worlds. The hard problem of "what it is like to be" a human – the problem David Chalmers identifies as that of subjective experience – from a Jamesean perspective is related to other problems concerning the biological and social limits of individual agency: How does the idea that motivates us to act come into being? One other novel idea of James's is that ideas shape our physiological response (Strawson 2008). By cultivating a habit of acting on ideas that may be individually repellant, but which are ethically and morally beneficial overall, we as a species can evolve cognitively. From the perspective of James's moral sensibility, then, the effort of attention itself is its own reward in the context of cultivating a conscious habit of attention.[16]

James's emphasis on this "moral" component to volition makes his ideas strikingly novel amidst experimental psychologists' efforts to understand psychology as merely involuntary physiological response to external stimuli, as is the case with James's contemporary, the psychophysicist E.B. Titchner of Cornell University.[17] James could not abide such reductionist interpretations of mind-brain dynamics to psychophysical formulae, despite his early training in the experimental psychology laboratories of Wilhelm Wundt and Herman von Helmholtz in Germany. He took issue, especially, with the founder of "psychophysics" Gustav Fechner's experimental correlation of "physiological bodily processes" with "immediately accompanying psychical events."[18] In his 1876 essay, "The Teaching of Philosophy in Our Colleges," James dismissed Fechner's psychophysical formula, writing, "It is more than doubtful whether Fechner's 'psychophysic law' (that sensation is proportional to the logarithm of its stimulus) is of any great psychological importance."[19] Later, in *Principles of Psychology* (1890), James dismisses psychophysics and in his first Gifford lecture on "Religion and Neurology," later published as *The Varieties of Religious Experience* (1902), he classes psychophysics as a branch of "medical materialism."[20] From James's perspective, psychophysics, concerned as it is with correlating brain states with finite sensory experience, leaves no room for the conscious self as an active selecting agency. Moreover, James argued, "Modern psychology, finding definite psycho-physical connections to hold good, assumes as a convenient hypothesis that the dependence of mental states upon bodily conditions must be thorough-going and complete."[21] Though this lecture seems to portray psychophysics as a pernicious validation of "medical materialism," James, in his later

---

[16] These ideas are expanded in his chapter on "Habit" in *The Principles of Psychology*. In 1887, the chapter appeared as an essay in *Popular Science Monthly*.

[17] Bjork 1983, pp. 81–83.

[18] Marshall 1982, p. 80. See also, Marshall 1974.

[19] qtd. in Marshall 1982.

[20] James 1990, p. 21.

[21] Ibid., p. 21.

writings, did not entirely relinquish Fechner's recognition of the interdependence of the psychical and the physical. James appropriated aspects of Fechner's metaphysical writings that comported with evolving ideas about the nature of matter coming into view as a result of the new physics. Probed to its minutest atomic structural elements, matter, like the mind itself, seemed in a constant state of flux.

James's concept of consciousness as a "theatre of simultaneous possibility" anticipates this new understanding of matter as permeable and discontinuous, based on the varieties of selective choice available to consciousness. As a consequence, consciousness itself is multitudinous. At the end of his long chapter in *Principles* on the "Self," in which James anatomizes what we think of as a singular "self" into an aggregate of material, social, spiritual, and psychological associations and responses to one's environment, he maintains that this "me is an empirical aggregate of things objectively known. The *I* which knows them cannot itself be an aggregate; neither for psychological purposes need it be considered to be an unchanging metaphysical entity like the Soul, or a principle like the pure Ego, viewed as 'out of time.'"[22] Rejecting both the Kantian notion of a transcendental ego and Cartesian dualism, together with the reductionist formulation of self as physiological brain states, James argues instead that, "It is a *Thought*, at each moment different from that of the last moment, but *appropriative* of the latter, together with all that the latter called its own." In other words, the "self," is a constantly shifting entity that is always already coeval with "the existence of passing thoughts or states of mind." Linking thought to the brain, James goes on to write that, "The same brain may subserve many conscious selves, either alternate or coexisting; but by what modifications in its action, or whether ultra-cerebral conditions may intervene, are questions which cannot now be answered."[23] James's *Principles* stopped short of solving the hard problem because the "science" of psychology "must stop with the mere functional formula." James thus concluded that "*thought is itself the thinker*."[24] Although James's *Principles* attempted to eschew metaphysics, his later works, which were directly informed by his psychical research in the decade before and after his writing on the *Principles*, pursued the relation of multiple conscious states to a "self" precisely by philosophically engaging such metaphysical questions as the survival of human consciousness after death and whether the brain itself could be understood as both a "transmissive" and "permissive" medium permitting information to leak in from "outside" through the permeable borders of the self. For this, and for his later pluralistic account of the "compounding" of consciousness, he drew upon his investigations of mediumistic trance and Fechner's more metaphysical speculations. James found validation for these metaphysical interests, I maintain, in the new physics and its grappling with new forms of matter: the equally mysterious rays and invisible subatomic particles.[25]

[22] James 1981, p. 379.

[23] Ibid., p. 379.

[24] Ibid., p. 379.

[25] McGinn's 2011 book on the philosophy of physics makes a similar argument about consciousness as a form of matter. As the new physics probed ever deeper into the hidden structures of material reality, "More and more types of physical reality had to be recognized, such as radio waves and X-rays, and the sheer versatility of matter became increasingly evident." Such discoveries gave rise

11.2 Evolving Forces: Psychical Research, Psychophysics, and the New Physics

James's involvement in the British Society for Psychical Research (SPR) began in 1885, following the death of his young son. His wife Alice had been invited to attend a mediumistic "sitting" held at the home of the young Boston medium, Mrs. Leonora Piper. Immediately convinced of the woman's authenticity, Alice brought James with her for subsequent sittings, thus launching what would become James's 25 year investigation of the medium. In 1909, nearing the end of his life, James confessed that he could draw no conclusive scientific meaning from Mrs. Piper's abilities, except for the existence of a persistent "will to personate," latent in all humans. James went on to connect this impersonating tendency to the evolution of thought itself, writing that such phenomena "are inwardly as incoherent as they are outwardly wayward and fitful. If they express anything, it is pure 'bosh,' pure discontinuity, accident, and disturbance, with no law apparent but to interrupt, and no purpose but to baffle. They seem like stray vestiges of that primordial irrationality, from which all our rationalities have been evolved."[26] Tellingly, James's language here suffuses evolutionary discourse with the very same language surrounding the new physics and the similarly "wayward and fitful" behaviors of invisible rays and subatomic energies. James's investigation of Mrs. Piper merely underscored the epistemological questions regarding James's philosophical "hard problem" – the relation of the "one and the many" – with which he had grappled for a lifetime. If James's earliest writings about consciousness and the self represented it as an "aggregate" of possible choices available to a selecting consciousness, his later writings attempt to understand how such a seemingly messy and disorganized self could nonetheless "know" its thoughts and communicate its intentions to others. If the "thinker" is the "thought," how would this self then address other similarly manifold other "selves" within a broader social community?

The manifold aspects of self that James encountered in his psychical investigations, as I have argued elsewhere, gave rise to his radically empiricist doctrine of pluralism, of possibility and choice, as the basis of consciousness.[27] James's championing of the discontinuity, indeterminacy, and flux suggested by his psychical research was anathema to psychologists invested in systematizing psychology by promoting experimental methods with reproducible, certain results, identifying psychophysical laws, and charting the brain's neurophysiologic coordinates for mental response. Indeed by the 1890s, James was declared the "nemesis" of all self-respecting psychologists invested in having psychology taken seriously as a scientific discipline. Those, like G. Stanley Hall, who had initially been his allies in founding the American Society for Psychical Research (ASPR) in 1884, fled its

to new "species of energy," such as "kinetic energy, chemical energy, gravitational energy, electromagnetic energy, nuclear energy" p. 176

[26] James 1986, p. 369.

[27] Hawkins 2005, 2011.

ranks to launch the American Psychological Association in 1890, which would become the bastion of scientific respectability for the new field.[28]

Experimental psychology promoted by Hall at Clark University, James Rowland Angell at the University of Chicago, Hugo Münsterberg at Harvard University, and Edward Bradford Titchener at Oxford and Cornell Universities – and later canonized by Edwin G. Boring's monumental *History of Experimental Psychology* (1929) – displaced an older American tradition of introspective self-scrutiny going back to the Puritans. The reigning narrative set forth by Boring, of the American experimental psychologist's descent from a German laboratory tradition, served to occlude the significant role that the religiously schooled early American "mental philosophers" played in the development of American psychology; moreover, transcendentalist and Swedenborgian elements pervade James's revaluation of religious experience as a valid source of intuitive knowledge.[29] Something else, however, accounts for James's invigorated metaphysical speculations after 1896. As religious historian Catherine Albanese has shown, the unknown and mysterious new forces unleashed by the discoveries of modern physics breathed new life into these older, mystical traditions belonging to the early Americas.[30] Though James has been associated with this earlier American transcendentalist psychological tradition, we can also think of James as a philosopher of science, specifically a philosopher of physics, since his lifelong concerns were mainly with concepts of energy and force, space and time, probability and possibility.

James was an avid student of physics from the beginning of his scientific career. Furthermore, if we track James's major publications with the discoveries that distinguish modern physics from the mechanistic, Newtonian worldview, the influence of the new physics on James's thinking becomes clear. James developed his "transmission theory" of consciousness in 1897, in which he described the brain's transmissive function in terms of invisible "rays," a mere 2 years after Wilhelm Roentgen's discovery of the X-ray. In 1902, the year Marie and Pierre Curie discovered the invisible element, radium; James lectured on mystical experience and the "reality of the unseen," later published as *The Varieties of Religious Experience*. James's heavily annotated copy of mathematician and philosopher of science Karl Pearson's second edition of *The Grammar of Science* (1900), a text that influenced Einstein, highlights the ways in which James's thought had taken a relativistic turn. Indeed, the lectures comprising *A Pluralistic Universe*, forming the basis for his philosophy called "radical empiricism," were published in 1904, just 1 year before Einstein arrived at his special theory of relativity. What I want to suggest is that these new discoveries in physics gave James a lexicon for describing consciousness, and reality itself, as palpably physical yet luminously immaterial. Moreover, what the historian Henry Adams at the turn-of-the nineteenth-century described as the "supersensual" domains disclosed by modern physics gave James a conceptual underpinning for his *Pluralistic Universe*, a universe honoring novelty, discontinuity, and ceaseless change within subjective experience.

[28] Coon 2002, p. 129.

[29] Fuchs 2002, pp. 79–84; Taylor 1992, 1996, p. 182, 2002, 2003.

[30] Albanese 2007.

These transformations in our understanding of the underlying composition of material reality as discontinuous and rife with novelty made possible James's later embrace of Fechner's metaphysical writings, known collectively as his "day view." As Michael Heidelberger has shown, Fechner's psychophysics is inextricably bound up with his metaphysical ideas, which enabled him to develop the ideas that would form the basis for contemporary psychophysics. Fechner attempted to resolve mind-body problems by identifying common naturalistic and empirical principles underlying all forms of life. Like James, Fechner maintained an indeterministic worldview that affirmed the existence of individual autonomy within a hierarchy of broader encompassing systems. As with James's pragmatism, the concept of "belief," or *glauben*, formed an essential component of Fechner's philosophical and scientific *Weltanschauung*. For Fechner, empirical observations came first. Subsequently, these observations became the basis for his metaphysical and natural-philosophic speculations regarding the constitution of the universe and the nature of human perception. For Fechner, the psychical (or mental) and the physical (or material) were different modalities of experience. Like the opposing sides of a single coin, the psychical and the physical were functionally parallel; they operated simultaneously, yet maintained an interdependency that was not linked by causality.[31] What has been termed Fechner's "double-aspect" view of the psychical and the physical, first described in his metaphysical work *Zend-Avesta oder über die Dinge des Himmels und des Jensits. Vom Standpunkt der Naturebetrachtung* ["On the Things of Heaven and the Afterlife: From the Standpoint of Meditating on Nature"] (1851), postulated a functional relationship between human experience and perception. His subsequent work, *Elements of Psychophysics* (1860) [*Elemente der Psychophysik*] delineated a mathematical means (later refined by Weber) of expressing this relationship that influenced mathematicians, such as James's friend Charles Sanders Peirce, in the USA, and the founders of German experimental psychology, Ernst Mach and Wilhelm Wundt, whose Leipzig laboratory was the training ground for a generation of American experimental psychologists who followed James.

In his lecture "On Human Immortality" (1898/1992), James made the German physicist his intellectual ally in exploring the most intractable of problems facing the mind sciences: the mind's relation to the brain and that of consciousness to human embodiment. James was most attracted to Fechner's metaphysical ideas, set forth in such works as *Little Book of Life After Death* [*Das Büchlein vom Leben nach dem Tode*] (1835; to which James wrote the introduction to the 1904 English translation) and *Zend-Avesta*, subtitled, "On the Things of Heaven and the Afterlife: From the Standpoint of Meditating on Nature" (1851). These metaphysical works lay the groundwork for Fechner's important 1861 *Elements of Psychophysics*,[32] and influenced James's radical empiricism, a philosophy that promotes a theoretical middle ground between strictly materialist and strictly metaphysical means of addressing the mind–brain problem.

[31] Marshall 1982; Heidelberger 2004.

[32] Marshall 1982; Heidelberger 2004.

From about 1905 until the time of his death in 1910, James was preoccupied by the German psychophysicist Gustav Fechner's panpsychic framework for understanding consciousness as constitutive of reality itself. Fechner's psychophysics stems from his development of a mind-body theory that attempts to account for the experienced unity of consciousness; Fechner maintained that the experience of inner states, or "self-phenomena," required coordination between "the inner aspect of a particular kind of unity and organization in a physical system." Fechner found it impossible to explain, a priori, how or why an individual's psychological experience of inner states appears unified. Instead, he argued for consciousness as a kind of "emergent property" that arises from the interaction of systems. In Part II of his *Zend Avesta*, Fechner provided the following hypothesis for how the interaction of two different systems might be experienced as one whole: "Juxtapose a system containing five points with another system containing five points, such that each system perceives the total connection of all its points to be one whole, in such a way that variations in the number and arrangements of the points create merely varying intensities and qualities of simple sensation. Now, one system is not linked to the other system in the same way that each is connected internally, for we do assume that they are two different systems. Thus the interconnectedness of another system will not be felt as strongly as one's own interconnectedness, but rather, we are affected by each point of the other as if it were singular."[33] In other words, two systems might overlap in such a way as to produce a sensation of wholeness that does not negate the separateness of each system. Fechner posits a unity that is not simultaneously a totality. Fechner's model attempts to escape Cartesian dualism, together with Hegelian and Kantian monistic idealism, and allows for individual particularity and autonomy within a larger naturalistic unity. This was a fitting model for James, who throughout his lifetime grappled with the problem of the "one and the many," the title he wanted to give to the systematic work of philosophy he left unpublished at the time of his death.

James's rhetorical stance in "On Human Immortality: Two Supposed Objections to the Doctrine" (1898/1900) was a strategic one; as a scientist tasked with making psychology into a respected science, he could not risk throwing it back into a mire of metaphysical speculation. Therefore, James asked his audience to take as gospel "the great psycho-physiological formula: Thought is a function of the brain."[34] If we take this formulation as a given, James asked, "Does this doctrine logically compel us to disbelieve in immortality?" James then bases the rest of his lecture on a philosophical thought experiment in which he uses the concept of immortality, or the survival of human consciousness beyond bodily death, to hypothesize a possible structure of consciousness in relation to the human brain quite apart from functional dependency. In place of the production theory, James argued for the brain's "permissive" or "transmissive" potential, in which the brain acts as a filter to information coming from outside.

[33] Heidelberger 2004, pp. 103–104.

[34] James 1992, p. 1104.

Describing the brain as analogous to a "prism, or a refracting lens," which transmits light, or to a pipe organ, through which air produces sounds, but is not itself "engendered in the organ," James argued that "mind is not generated by the brain but instead focused, limited, and constrained by it."[35] For James, the phenomenon we think of as "mind," cognition, or mental awareness, is a consequence of the brain's behaving as a kind of receiving station to "the genuine matter of reality" transmitted by the environment.[36] There is a peculiarly Jamesean legerdemain in not naming the substance of this reality, except through suggestive metaphors: invisible light, the trajectory of an arrow shooting through air, air passing through the apparatus of a pipe organ or, more poetically as a "white radiance." Consciousness, James maintained, was a "sphere of being" that is "continuous" with "that more real world."[37] Of what invisible substance "genuine" reality was composed, James would leave to others to discover.[38]

James's "transmission theory" was indebted to the ideas of at least one other key individual: Frederic H. Myers, founder of the British Society for Psychical Research. James's transmission theory was modeled in part on Fechner's "conception of a fluctuating psychophysical threshold," while his notion of the self as an entity that contains a plurality of possible mental states and secondary "personalities," drew upon Myers's concept of a "subliminal" or "supraliminal" Self – an entity that encompasses a field of broader awareness coexistent with a subject's more narrow sense of a coherent self, but that is not necessarily restricted by or even known to that primary self. In developing his "transmission theory," James had refined Myers's theory of the Subliminal Self by being the first to explicitly link "notions of transmission and filtering with the brain" (through the metaphor of the "prism" through which light passes), only to come out on "the other side filtered, reduced, focused, redirected, or otherwise altered in some systematic fashion."[39]

Just a decade later, James would be confounded by an experience that would seem to validate his hypothesis concerning the brain's transmissive potential. On January 28, 1909 James presented his "Report on Mrs. Piper's Hodgson-Control" (1909) for the General Meeting of the Society for Psychical Research (SPR). The report concerned one Richard Hodgson, secretary for the American SPR, as well as an avowed agnostic and a dogged debunker of "spiritualist" phenomena and persistent investigator of the Boston medium Leonora Piper. So vexed was he by his inability to expose her as a fraud, Hodgson once jokingly vowed that if he died, he would return to haunt Mrs. Piper as her control. As if in fulfillment of this prophecy, after Hodgson's sudden death of a heart attack, he reportedly began making

[35] Ibid., pp. 1109–1110; Kelly et al. 2007, p. 29.

[36] James 1992, p. 1111.

[37] Ibid.

[38] As James himself acknowledged, philosophers Immanuel Kant and F.C.S. Schiller made similar arguments. Kant, for example, maintained that the body restricts the intellectual function of the brain, which only comes into full flower after death. Schiller similarly argued that matter restricts "the consciousness which it encases." Ibid., p. 1119, n9.

[39] Kelly et al. 2007, p. 606.

appearances as one of Mrs. Piper's spirit "controls." A wholesale investigation ensued, and James, having befriended Hodgson during his own investigations of Mrs. Piper, was at the center of it. For many months James followed in vain the baroque narratives of Mrs. Piper's so-called "Hodgson-control," finding them not only tiresome and trite, but annoyingly inconclusive regarding any insight they provided to the biological or metaphysical aspects of trance states. James's report, however, provides an interesting coda to our exploration of this transmission theory and the compounding of consciousness, for in the report James invokes Fechner as an "aid" to his "imagination."[40]

In order to account for facts provided by the Hodgson-control that were verifiably true, but apparently unknown to the conscious Mrs. Piper, James looked to Fechner's *Zend-Avesta*, citing his idea that "mental and physical life run parallel, all memory-processes being, according to him, co-ordinated with material processes. If an act of yours is to be consciously remembered hereafter, it must leave traces on the material universe such that when the *traced parts of the said universe systematically enter into activity together* the act is consciously recalled. During your life the traces are mainly in your brain; but after your death, since your brain is gone, they exist in the shape of all the records of your actions which the outer world stores up as the effects, immediate or remote, thereof, the cosmos being in some degree, however slight, made structurally different by every act of ours that takes place in it."[41] In other words, what James suggests, vis-à-vis Fechner, is that while we are living our actions leave a trace or residue in physical form, which can later can be picked up and revivified like a dormant current of life suddenly awakened by another's sympathetic response. James went on to postulate that our brains are like so many different "Marconi-stations."[42]

11.3 James's Inheritors: Neurophenomenology and Physics

It would be all too easy to dismiss James's report, together with his "transmission" theory of consciousness, as metaphysical clap-trap were it not for the fact that several recent mind philosophers and physicists have arrived at a similar account of mind-brain dynamics. Biomedical engineer Paul Nuñez's model of the consciousness as a "nested hierarchy" for processing information from the environment posits the mind and environment as co-dynamic, mutually constitutive entities in ways that resemble James's model of the brain as a "filter."[43] Making no reference to James's transmission theory, Nuñez then goes on to posit "a highly speculative" account of consciousness that is nonetheless dramatically similar to that of James when he

[40] James 1986, p. 357; for accounts of the experiences leading up to James's report, see Richardson 2007, pp. 478–479; for Hodgson's reputed vow, see Simon 1998, pp. 319–320.

[41] Ibid., p. 358, emphasis in original.

[42] Ibid., p. 359.

[43] Nuñez 2010, p. 11.

describes how "whole brains or special parts of brains might behave like antenna systems sensitive to an unknown physical field or other entity that, for want of a better name, may be called Mind."[44] In this way, James's account of the brain's "transmissive" properties paved the way for more contemporary accounts assigning the mind–brain specific temporal–spatial dimensions and a hierarchical structure.

James's views expressed in his report on the "Hodgson control" are also strikingly consistent with recent attempts to understand consciousness as representing the same "hard problems" that plague quantum systems. Quantum physicist Thomas Filk and mind philosopher Albrecht von Müller maintain that the irreducibly unpredictable nature of consciousness and the ensuing "indeterminacy" of possible outcomes make it "categorically" similar to quantum physics. "Both consciousness and quantum theory," they observe, "refer to the *status nascendi* aspect of reality."[45] Consciousness is coterminous with all the latent possibilities contained in our experience of the present. What we think of retrospectively as "fact" resolves itself out of these latent, or nascent, possibilities. Filk and Müller subsequently define fact as "the lasting results of an 'event', i.e., the traces or imprints of a previous event has left in the present state of the universe. These traces can be memories (imprints in the neural structure of our brain), books, pictures, fossils, 'documents' … or other forms of recording."[46] "Fact," as defined by Filk and Müller comes to resemble what James understood as the trace, or residue, of past events. James's correspondent in his later years, the French philosopher Henri Bergson, in *Matter and Memory* (1896), which James had read, similarly argued that "every movement leaves traces that continue to affect all subsequent physical or mental processes."[47] In sum, mental "events" are selected from the "theater" of simultaneous possibilities located in the present, or as Edelman would have it, the "remembered present" of conscious experience.[48]

From a less metaphysical point of view, other contemporary mind scientists reinforce the validity of James's phenomenal investment in mind as a dynamical system (a theater of possibility) in which consciousness and material reality are interdependent. For example, cognitive psychologist Arnold Trehub's theoretical model of the "retinoid system," which allows the brain to produce the phenomenal time-space that we come to associate with an individual ego or "I," has explained that "Our cognitive brain is especially endowed with neuronal mechanisms that can model within their biological structures all conceivable worlds, as well as the world we directly perceive or know to exist. External expressions of an unbounded diversity of brain-created models constitute the arts and sciences and all the artifacts and enterprises of human society."[49] Although consciousness gives up its ghost, so to speak, in this model, it nonetheless retains elements of Jamesean possibility.

[44] Ibid., p. 274.

[45] Filk and Müller 2008, p. 63; See Stapp 1995/1997, 2007.

[46] Filk and Müller 2008, p. 63.

[47] Kern 1983, p. 41.

[48] Edelman 2004, p. 133.

[49] Trehub 2007, p. 310.

The provisional, indeterminate and spontaneous aspects of consciousness that James associated with ecstatic religious experience, exceptional mental states, and the intuitive leaps of genius that James associated with the uniqueness of the individual self is retained in Trehub's model. Though he does not acknowledge the metaphor's Jamesean origins, it is worth quoting Trehub at length on his extended metaphor for consciousness as a "theatre":

> What happens on the stage of retinoid space is never a finished production. Metaphorically, it is always a work in progress with constantly changing scenes and cast, directed partly by the current environment and partly by the sensory and evaluative neuronal mechanisms in the 'dark' of the theater as they respond to the brain's changing motives and tactics in its encounters with the natural world. There are, however, two constant aspects to the production on the stage of consciousness–one is the space-time envelope of the retinoid stage itself, and the other is the ubiquitous participating self. So, in a metaphorical sense, the retinoid system is the 'illuminated' stage of consciousness, the evaluative/executive neuronal mechanisms (the audience) are off-stage. The other supporting adaptive processes of the brain and its body might be loosely thought of as the stage hands.[50]

The "audience," he notes, is represented by "the outputs of all evaluative brain mechanisms (all members of the 'audience') are implicit aspects of a unitary self."[51] Trehub's view does not posit a homunculus hidden in the recesses of the corpus callosum; rather, it represents the field of awareness and possibility produced by the retinoid system of dynamically integrated neuronal synapses related to our perceptual awareness and our phenomenal experience of reality.

In his later writings on the "compounding of consciousness," James treads a step further than modern neuroscience by addressing what happens when the "audience" is in fact other consciousnesses. Here, James gets at the larger sociological dimensions of the "hard problem" of reconciling individual minds to larger communities of individuals. When James later discussed the "compounding of consciousness" in the fifth chapter of his *Pluralistic Universe*, he extended the idea of the brain's transmissive properties to include that of other consciousnesses, writing,

> My present field of consciousness is a centre surrounded by a fringe that shades insensibly into a subconscious more… The centre works in one way while the margins work in another, and presently overpower the centre and are central themselves. What we conceptually identify ourselves with and say we are thinking of at any time is the centre; but our full self is the whole field, with all those indefinitely radiating subconscious possibilities of increase that we can only feel without conceiving, and can hardly begin to analyze.[52]

In writing this, James was thinking expressly of Fechner's psychophysical threshold, now known as the Weber–Fechner law, postulating that "consciousness" is the threshold at which subjective perception and subjective sensation coincide. James was less interested in the mathematical formulation for this law than he was in the assigning of temporal–spatial movement to consciousness. These "movements," as James would write in his introduction to the English translation of

[50] Ibid., 329.

[51] Ibid.

[52] James 1977, p. 130.

Fechner's *Little Book of Life After Death*, "can be superimposed and compounded, the smaller on the greater, as wavelets upon waves. This is as true in the mental as in the physical sphere. Speaking psychologically, we may say that a general wave of consciousness rises out of a subconscious background, and that certain portions of it catch the emphasis, as wavelets catch the light… On the physical side we say that the brain-processes that corresponded to it altered permanently the future mode of action of the brain."[53] What James was arguing – drawing upon Fechner's model of the threshold of consciousness as a sinusoidal wave – is richly suggestive of dynamical systems. James's point of view similarly accords with that of phenomenologist Evan Thompson, who collaborated with the late Francisco Varela to write *Mind in Life* (2007). In this phenomenological account of neurophysiological processes, Thompson understands "dynamical systems" as "a collection of related entities or processes that stands out from a background as a single whole, as some observer sees and conceptualizes things."[54] The solar system is one such example, but James's transmission theory offers the example of the social environment, in which one consciousness coexists among many others. In a very real sense, the compounding of consciousness suggests the co-penetration of individual consciousnesses within ever larger and interpenetrating systems.

This idea that consciousnesses themselves co-penetrate is made explicit in an even earlier passage, from the first lecture in *A Pluralistic Universe*. In distinguishing monism from his philosophical pluralism, James writes: "My thoughts animate and actuate this very body which you see and hear, and thereby influence your thoughts. The dynamic current somehow does get from me to you, however numerous the intermediary conductors may have to be. Distinctions may be insulators in logic as much as they like, but in life distinct things can and do commune together every moment."[55] The world of *A Pluralistic Universe* is just such a dynamical system comprising a world of interconnecting relations, of "complexity-in-unity" enveloped by a surrounding "earth-consciousness."[56] And here we finally arrive at the panpsychic view James adopted later in life and attributed to Fechner (Lamberth 1997 p. 250).

This philosophical position of James's strongly accords with the contemporary neuroscientific theory of "dynamic co-emergence," held by Thompson and Varela, in which living and mental processes are understood as "unities or structured wholes rather than simply as multiplicities of events external to each other, bound together by efficient causal relations."[57] In phenomenological terms, this means revising our understanding of nature as "not pure exteriority," but rather as possessing "its own interiority." Thompson is careful to distinguish this perspective from "metaphysical idealism," the argument for a "preexistent consciousness." Instead, it implies a "transcendental orientation" by which we understand that "the world is never given to us as a brute fact detachable from our conceptual frameworks. Rather, it shows

[53] James 1904, p. xv.

[54] Thompson 2007, p. 39.

[55] James 1977, pp. 115–116.

[56] James 1977, p. 73, 1909, 1910.

[57] Thompson 2007, p. 67.

up in all the describable ways it does thanks to the structure of our subjectivity and our intentional activities."[58] James would understand this in terms of an inherent intimacy of relations between the self and the world with which the self engages. Consciousness itself is "transcendent," in Thompson's terms, in part because, as he says, it "is always already presupposed as an invariant condition of possibility for the disclosure of any object [;] there is no way to step outside, as it were, of experiencing subjectivity, so as to effect a one-to-one mapping of it onto an external reality purged of any and all subjectivity."[59] Consciousness seems defined then by some variable movement or change in time that is perceived differently in relation to one's location in time and space, and that also depends upon one's particular role and orientation toward the experiment, that is, whether one is experiencing mental phenomena as a subject in an experiment or as the witnessing and recording observer. In light of Thompson's phenomenological orientation toward the mind–brain conundrum, it is this intersubjective dimension that becomes most salient to the future of contemporary mind–brain research.

Picking up where James left off, with the notion that subjective experience itself is the basis for decision-making and that the facts of subjective experience compose the intricate web of social "facts" that humans negotiate, Chalmers writes that the "really hard problem of consciousness is the problem of *experience*."[60] The hard problem gives rise to a series of hard questions about the relationship between what we experience as a subjective self and the various operations of the physical brain: How does our complex sensory experience of the environment become interiorized, coterminous in our own thoughts with a perceptible "self"? The more pressing question, for Chalmers, is not how consciousness works, but, more precisely, *why*: "How can we explain why there is something it is like to entertain a mental image or to experience an emotion?".[61] What makes the hard problem hard is that it is not about "the performance of functions."[62] Among mind scientists grappling with the "hard problem" there seems to be a consensus forming around first, the idea that, "a unified science of consciousness" requires anchoring in "the biological sciences,"[63] and second, that there is an essential complementarity between the subjective mind and the physical brain.[64] To these two premises, I believe James would heartily have agreed, for he too was concerned to identify a naturalistic account for subjective mental phenomena – ecstatic experience, hallucinations, and mediumistic trance.[65]

---

[58] Ibid., p. 82.

[59] Ibid., p. 87.

[60] Chalmers 2010, p. 5.

[61] Ibid.

[62] Ibid., p. 6.

[63] Revonsuo 2009, p. 3; Fingelkurts et al. 2010; Kelly et al. 2007.

[64] Pauli 1994, p. 260; Fingelkurts et al. 2010.

[65] See McGinn on consciousness as a form of matter: He writes, "An electromagnetic field is a type of material reality, and so is consciousness. Alternatively, consciousness is one form of energy, along with kinetic energy or electrical energy. If this hypothesis is true, then consciousness is

The larger problem for James, however, was how to overcome the epistemological and methodological problems related to identifying how interior states of consciousness correspond to physiological processes, an effort that necessarily would rely on accurate self-reporting and careful observation by investigators. From James's perspective, investigators would have to be self-observers as well, attuned to the ways in which their own biases might influence and predetermine results. In today's terms, an investigator serves as a witness to the self-reporting of phenomena, in tandem with seemingly "objective" visualizing technologies, such as EEG, or other means of visually representing internal cognitive processes. But these are merely descriptive of processes and fail to address the more complex actions that lead to individual and collective decisions, including the decision-making of researchers themselves (see Jack and Roepstorff 2002). To address this problem, investigators have called for a more "phenomenologically oriented psychology," one that focuses on "the phenomenology of the science-making process itself, and the experimenter as the new confounding variable in the conduct of experiments."[66] The study of what Varela termed "neurophenomenology," writes Taylor, would address an epistemological divide between the neuroscientific and philosophy of mind approaches. The problem is that the brain is physical, while the "mind" is impossible to locate; to be more accurate, "the mind is a metaphor for experience."[67] And this experience, for James, was riddled with inconsistencies. One of the hardest problems of all, as James noted, is that there is no transcendently true experience that holds for all individuals at all times.

This essay's title taken from James's metaphor for consciousness as a "theatre of simultaneous possibilities" highlights a significant point of contact between James and contemporary neuroscientific responses to the hard problem. From this Jamesean point of view, consciousness is irreducibly complex, a dynamic system, in which both automatic, and seemingly "willed," processes of neural selection produce indeterminate outcomes. What we think of as the self or "mind" is an emergent property as a consequence of the dynamic interplay of internal neuronal processes that produce our retinal imaging of a spatio-temporally bounded "external" world that corresponds to our experience of being an individuated "self" that lives, moves, and volitionally selects from many possibilities within this phenomenal world. Neuroscientists, physicists, and philosophers of mind each describe the neural mechanisms and their dynamically complex interactions responsible for producing this phenomenal space variously as the "phenomenal transform"; the "Retinoid Model"; or the brain's "Operational Architectonics."[68] All of these, I suggest, are indebted to James's recognition of consciousness as a complex system, whose eventual understanding will require the concerted efforts of an equally complex, disciplinarily diverse, and similarly dynamically-integrated scientific community.

material after all – though not in the Cartesian sense. The general conception I am working with is that matter/energy is the underlying substance of the universe, and it may ultimately be unitary, but it can take widely different forms – with consciousness as just one of them" 2011, p.178

[66] Taylor 2010, p. 411; Wallace 2007, p. 105.

[67] Varela 1997; Taylor 2010, pp. 421–422.

[68] Edelman 2004, p. 76; Trehub 2007; Fingelkurts et al. 2010, 2011, 2012.

11.4 Conclusion: The Social Force Fields of Consciousness

The first to understand consciousness as a perceptual "field" of awareness, composed of a center and an outer "fringe" or "margin," James early on associated the hard problem of consciousness – its seeming immateriality, the impossibility of measuring it, the difficulty of identifying consistent and immutable laws – with the new physics. More specifically, James linked the concept of "force" with the concepts of "will" and "volition" – in this, he anticipated Walter J. Freeman's view that "consciousness is not merely 'like' a force: it *is* a field of force that can be understood in the same ways that we understand all other fields of force (and energy) within which we, through our bodies, are immersed, and which we, through our bodies, comprehend in accordance with the known laws of physics."[69] Its stream, argues Freeman, is "cinematographic rather than continuous" – a series of frames – and consciousness's "prime role is not to make decisions but to delay and defer action and thereby minimize premature commitment of limited resources."[70] What I have attempted here is, first, to recuperate James's place as a pioneer of the hard problem related to the subjective nature of experience and volition, and second, to reclaim early psychophysics as significant for James's theory of experience as the dynamic co-emergence of consciousness and the social realm. In other words, from a Jamesean perspective, the hard problem is a biosocial phenomenon as much as it is a neurophysiological one. Everything that researchers have discovered about the human brain reveals that we are pattern-seeking, meaning-making animals. We do so individually, and, moreover, in groups.

The question that arises, then, is in what ways does addressing the "hard problem" matter not just for neuroscience and philosophy of mind, psychology, and the natural and social sciences, but for humanistic inquiry overall? In the last two decades, the neuroscientific debate about consciousness has hinged on whether or not the "hard problem" itself actually exists and, if so, whether solving it is important.[71] There is a further aspect of the hard problem that contributes to its difficulty: What role, if any, do culture and the social realm play in influencing the evolution of individual minds or brains? As a late Victorian thinker, James devoted his psychology and philosophy to moral realms; his theory of consciousness had implications not only for the social and moral frameworks of his cultural moment, but beyond. James never abandoned the philosophical problem of the "one and the many," first represented as monism and pluralism, but in his later writings he elaborated this problem to include the relation of the individual mind to the masses. Like many modernist artists and intellectuals, James feared mass psychology. He found something dangerous in groupthink, while, like his godfather, the New England transcendentalist Ralph Waldo Emerson, he saw something ameliorative in the mind of individual geniuses who were the chief architects of new ideas that revitalize collective human thought.

[69] Freeman 2007, p. 1022.

[70] Ibid.

[71] Shear 1997.

In the decade before his death, James's thinking about consciousness became increasingly grounded in the tumultuous political and social events of his era: imperialism and racist violence against African-Americans. James's wariness of the mob and of groupthink in general reinforced his understanding of consciousness as an entity that was not merely mechanical; rather, the interactions of individual consciousnesses, with their potential for "compounding" individual thoughts into collective beliefs, potentially disclosed the trickier sociological questions regarding the production and development of individual and group beliefs and behaviors. Expressed in terms of James's philosophical pragmatism, the existence of "consciousness" as an entity or a property belonging to an individual was known by its effects in and upon the world and others. For James, there was a social dimension to the hard problem, which gave the pursuit of its resolution a special kind of urgency. His outspoken critiques of U.S. imperialism and racism, particularly its manifestation in the mob mentality of lynching, all result from his conviction that the individual expressions of volition could represent either a benevolent or a diabolical genius. The former could spur human cognitive evolution to greater moral heights, while the latter, conversely, could precipitate a slide toward moral degeneracy and social depravity. James's writings on the hard problem of consciousness arising from individual volition (attention, habit, will) led him indirectly into the investigation of the neuroscientific roots of public opinion.[72]

My and your experience is not just mine or yours alone, James would argue, but it is something that makes a potential impact on the world; my unique awareness can be shared with you and with others, and, in the process, acquire a kind of agency. Thought itself, in other words, becomes something tangible – a force – insofar as it has the capability of literally moving individual minds and bodies to collective social action. Under James's philosophical gaze, consciousness becomes something wildly potent. James was drawn to the work of such metaphysical thinkers as Gustav Fechner and Henri Bergson precisely because they shared his conviction that thought itself is a force. It leaves a trace, long after the individual agent has departed the world of the living, on human communities and collective memory.

Acknowledgment I am grateful to Andrew and Alexander Fingelkurts for commenting on an earlier draft of this essay.

References

Albanese CL (2007) A republic of mind and spirit: a cultural history of American metaphysical religion. Yale University Press, New Haven/London

Barnard GW (1997) Exploring unseen worlds: William James and the philosophy of mysticism. State University of New York Press, Albany

[72] Between 1899 and 1904 James published anti-imperialist essays against U.S. occupation of the Philippines. His anti-lynching essays appear in 1903, on the heels of James's many reviews of books dealing with crowd psychology and mental hygiene. James 1987.

Bjork DW (1983) The compromised scientist: William James in the development of American psychology. Columbia University Press, New York
Bordogna F (2008) William James at the boundaries: philosophy, science, and the geography of knowledge. University of Chicago Press, Chicago/London
Chalmers DJ (2010) The character of consciousness. Oxford University Press, Oxford
Coon DJ (2002) Testing the limits of sense and science: American experimental psychologists combat spiritualism, 1880–1920. In: Pickren WE, Dewsbury DA (eds) Evolving perspectives on the history of psychology. American Psychological Association, Washington, DC, pp 121–139
Cotkin G (1990) William James: public philosopher. Johns Hopkins University Press, Baltimore
Crippen M (2010) William James on belief: turning Darwinism against Empiricistic Skepticism. Trans Charles Peirce Soc 46(3):477–502
Edelman G (2004) Wider than the sky: the phenomenal gift of consciousness. Yale University Press, New Haven/London
Feinstein HM (1999) Becoming William James. Cornell University Press, Ithaca/London
Filk T, von Müller A (2008) Quantum physics and consciousness: the quest for a common conceptual foundation. Mind Matter 7(1):59–79
Fingelkurts AA, Fingelkurts AA, Neves CFH (2010) Natural world physical, brain operational, and mind phenomenal space-time. Phys Life Rev 7(2010):195–249
Fingelkurts AA et al (2011) Persistent operational synchrony within brain default-mode network and self-processing operations in healthy subjects. Brain Cogn 75:79–90
Fingelkurts AA, Fingelkurts AA, Bagnato S, Boccagni C, Galardi G (2012) Toward operational architectonics of consciousness: basic evidence from patients with severe cerebral injuries. Cogn Process 13:111–131
Freeman WJ (2007) Indirect biological measures of consciousness from field studies of brains as dynamical systems. Neural Netw 20:1021–1031
Fuchs AH (2002) Contributions of American mental philosophers to psychology in the United States. In: Pickren WE, Dewsbury DA (eds) Evolving perspectives on the history of psychology. American Psychological Association, Washington, DC, pp 79–99
Hawkins SL (2005) The science of superstition: Gertrude Stein, William James, and the formation of belief. Mod Fict Stud 51(1):60–87
Hawkins SL (2011) William James, Gustav Fechner, and early psychophysics. Front Fractal Physiol 2:1–10
Heidelberger M (2004) Gustav Theodor Fechner and his psychophysical worldview (trans: Klohr C). University of Pittsburgh Press, Pittsburgh
Jack AI, Roepstorff A (2002) Introspection and cognitive brain mapping: from stimulus-response to script-report. Trends Cogn Sci 6(8):333–339
James W (1879/1983) Are we automata? Mind 4(January 1879):1–22. In: Burkhardt FH, Bowers F, Skrupskelis IK (eds) Essays in psychology. The works of William James. Harvard University Press, Cambridge, MA/London, pp 38–61
James W (1888/1983) What the will effects. Scribner’s Magazine 3(February 1888):240–250. In: Burkhardt FH, Bowers F, Skrupskelis IK (eds) Essays in psychology. The works of William James. Harvard University Press, Cambridge, MA/London, pp 216–234
James W (1904) Introduction. In: Fechner G (ed) The little book of life after death (trans: Wadsworth MC). Little, Brown, and Company, Boston, pp vii–xix
James W (1909) The doctrine of the earth-soul and of beings intermediate between man and God; an account of the philosophy of G.T. Fechner. Hibbert J: A Q Rev Relig Theol Philos 7(2):278–291
James W (1910) A suggestion about mysticism. J Philos Psychol Sci Method 7(4):85–92
James W (1922) Essays in radical empiricism. Longmans, Green and Co., New York (Original 1904)
James W (1977) A pluralistic universe, vol 4, The works of William James. Burkhardt F, Bowers F (eds). Harvard University Press, Cambridge, MA (Original 1909)
James W (1981) The principles of psychology. The works of William James. Burkhardt F, Bowers F, Skrupskelis I (eds). Harvard University Press, Cambridge, MA/London (Original 1890)
James W (1986) Essays in psychical research. Harvard University Press, Cambridge
James W (1987) Essays, comments, and reviews. Burkhardt FH, Bowers F, Skrupskelis IK (eds). Harvard University Press, Cambridge, MA/London

James W (1990) The varieties of religious experience. Kuklick B (ed). The Library of America, New York

James W (1992) On human immortality. In William James: writings 1878–1899. The Library of America, New York, pp 1100–1127 (Original 1898)

Kelly EF et al (2007) Irreducible mind: toward a psychology for the 21st century. Rowman and Littlefield, New York

Kern S (1983) The culture of time and space: 1880–1918. Harvard University Press, Cambridge, MA

Lamberth DC (1997) Interpreting the universe after a social analogy: intimacy, panpsychism, and a finite God in a pluralistic universe. In: Putnam RA (ed) The Cambridge companion to William James. Cambridge University Press, Cambridge, pp 237–259

Leary DE (2002) William James and the art of human understanding. In: Pickren WE, Dewsbury DA (eds) Evolving perspectives on the history of psychology. American Psychological Association, Washington, DC, pp 101–120

Marshall M (1974) William James, Gustav Fechner, and the question of dogs and cats in the library. J Hist Behav Sci 10:304–312

Marshall M (1982) Physics, metaphysics, and Fechner's psychophysics. In: Woodward WR, Ash MG (eds) The problematic science: psychology in nineteenth-century thought. Praeger, New York, pp 65–87

McGinn C (2011) Basic structures of reality: essays in metaphysics. Oxford University Press, New York

Nuñez PL (2010) Brain, mind, and the structure of reality. Oxford University Press, Oxford

Pauli W (1994) Writings on physics and philosophy, Enz CP, von Meyenn K (eds). Springer, Berlin

Revonsuo A (2009) Inner presence: consciousness as a biological phenomenon. The MIT Press, Cambridge, MA/London

Richardson RD (2007) William James: in the maelstrom of American modernism. Mariner, New York

Schwartz JM, Begley S (2003) The mind and the brain: neuroplasticity and the power of mental force. Harper Perennial, New York

Shear J (1997) The hard problem: closing the empirical gap. In: Explaining consciousness: the 'Hard Problem'. The MIT Press, Cambridge, MA/London, pp 359–375

Simon L (1998) Genuine reality: a life of William James. Harcourt Brace, New York

Skrupskelis IK (1995) James's conception of psychology as a natural science. Hist Hum Sci 8(1):73–89

Stapp HP (1995/1997) The hard problem: a quantum approach. In: Shear J (ed) Explaining consciousness: the 'hard problem'. The MIT Press, Cambridge, MA/London, pp 197–215

Stapp HP (2007) Mindful universe: quantum mechanics and the participating observer. Springer, Berlin

Strawson G (2008) Real materialism and other essays. Oxford University Press, New York

Taylor E (1992) The case for a uniquely American Jamesian tradition in psychology. In: Donnelly ME (ed) Reinterpreting the legacy of William James. American Psychological Association, Washington, DC, pp 3–28

Taylor E (1996) William James on consciousness beyond the margin. Princeton University Press., Princeton

Taylor E (2002) William James and depth psychology. In: Michel F (ed) The varieties of religious experience: centenary essays. Imprint Academic, Charlottesville, pp 11–36

Taylor E (2003) A perfect correlation between mind and brain: William James's *Varieties* and the contemporary field of mind/body medicine. J Specul Philos New Ser 17(1):40–52

Taylor E (2010) William James and the humanistic implications of the neuroscience revolution: an outrageous hypothesis. J Humanist Psychol 50:410–429

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

Trehub A (2007) Space, self, and the theater of consciousness. Conscious Cogn 16:310–330

Varela FJ (1997) Neurophenomenology: a methodological remedy for the hard problem. In: Jonathan S (ed) Explaining consciousness: the 'hard problem'. The MIT Press, Cambridge, MA/London, pp 337–357

Wallace BA (2007) Hidden dimensions: the unification of physics and consciousness. Columbia University Press, New York

Chapter 12
The Enigmatic Deciphering of the Neuronal Code of Word Meaning

Andrew C. Papanicolaou

Mind and body are one and the same individual
which is conceived now under the attribute of thought,
and now under the attribute of extension.

Baruch Spinoza

.....in respect to the form of appearances, much may be said .....
whilst of the thing in itself, which may lie at the foundation
of these appearances, it is impossible to say anything.

Immanuel Kant

In the final year of the decade of the brain, an important article, titled "Words in the Brain's Language" made its appearance in the Journal of "Behavioral and Brain Sciences".[1] It develops the theoretical ramifications of Hebb's notion of cell assemblies[2] for understanding first, how neuronal codes of word meanings are established; second, where the assemblies are located and third, on the basis of what logic they are arranged on the cortical surface. In the critical discussion that accompanied this "target" article, the question inevitably arose as to how, precisely, these patterns of neuronal signals cause (give rise to, are transformed into, or constitute the basis of) the experienced meaning. The question exemplifies the "hard problems" that have been hounding psychology, philosophy, linguistics and systems neurosciences, throughout their respective histories and which surface whenever the natural sciences are called upon to account for

[1] Pulvermüller 1999.

[2] Hebb 1949.

A.C. Papanicolaou (✉)
Department of Pediatrics, Division of Clinical Neurosciences,
The University of Tennessee Health Science Center, Memphis, TN, USA
e-mail: apapanic@uthsc.edu

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_12,

mind, meaning and consciousness. To that question, the author of the target article responded as follows:

> …. I must confess, I am unable to answer…… What makes an assembly a concept? Why do I consciously experience a certain association when reading the word "mouse"? I believe that these questions can not be answered, and popular statements that consciousness starts at 5 μV (as Libet's 1985 results suggest) or is apparent in 40 Hz activation (Crick and Koch 1990; Koch and Crick 1994) are somewhat unsatisfactory, because the questions can be iterated: why should strong electrocortical potentials and high frequency spatio-temporal patterns make me experience consciously?[3]

And he concluded:

> The activation of large (and strongly linked?) cortical neuron populations is the physical basis of consciousness. Further questions will probably lead to nothing but confusion.[4]

One would hardly be blamed for agreeing with the author's pessimistic sentiment: There is nothing, or next to nothing, of substance that modern scholarship has added to the speculation of the philosophers of the past, on this problem. Contemporary neuro-philosophy (see e.g. Churchland[5]) has only managed to add to these philosophical endeavors the dubious qualification that the prefix "neuro-" confers, and the apparently irrelevant to the main question, therefore confusing, specification of the various metrics of the activation of the neuronal assemblies like voltages and oscillation frequencies.

But the reason the modern no less than the ancient forays into this labyrinthine region have been unsuccessful, may have nothing to do with the scientific legitimacy of the questions, as the concluding remarks of the author appear to imply. Rather, it may have everything to do with the fact that the questions are stated in a way that constrains severely the type of answers they permit. Specifically, the questions do not demand an explanation of the relation between neuronal codes and experiences, whatever that relation may turn out to be but, instead, they demand an account as to how the codes cause experiences (having excluded thereby other possible forms the relation) – a demand, moreover, not based on evidence but on the tacit metaphysical assumption that either neuronal signals cause experiences directly, or that there is a mechanism in the brain that transforms them into experiences. As for the assumption, it has proved to be singularly unproductive, as evidenced by the fact that although countless models that rest on it have been developed over the centuries, they have all failed to generate a sufficiently persuasive answer to settle the issue to the satisfaction of most scientists and philosophers but also to the satisfaction of the public at large.[6]

[3] Pulvermüller 1999, p. 326.

[4] Ibid., p. 326.

[5] Churchland 1986, 2002a, b.

[6] That the issue has not been settled is amply attested by the unabated debate which has intensified in recent years see e.g. Pockett et al. (2009) and by the fact that the questions of whether and to what degree freedom of will (and of overt action) is conditioned by the physiology of the brain, are debated in thousands of courts of law, every day, throughout the world.

In this essay, I propose to demonstrate that whereas specific questions as to how neuronal codes cause meanings are in fact confusing and metaphysically prejudiced, the issue of the relation of neuronal codes to meanings is a perfectly reasonable subject of empirical inquiry; that far from breeding confusion it admits of a clear answer and that it does so not by means of metaphysical argument but by the consistent and exclusive application of the methods of the natural sciences. Before, however, it will be helpful to summarize the main points of the version of the cell assembly theory presented in that seminal article I referred to above, and to clarify the meaning of the word "meaning" as it was used in that article and will be used in this essay. The latter, I believe, is especially needed in view of the fact that whereas there is broad agreement as to what a neuron, a cell assembly, a voltage, or even a neuronal code might be, agreement as to what might be meant by the "meaning" of a word, is far from universal.

In this context, then, the word "meaning" stands for the private psychological phenomenon, that manifold experience we all have when we encounter known words in context or in isolation. The experience is multifaceted in that when we reflect on it retrospectively (which is the only way we can reflect on any experience) it is possible to discern some of its constituent elements. A similar analysis has in fact been formally implemented and the constituent elements of word meaning have been systematically classified over half a century ago by Osgood, Suci and Tennenbaum.[7] The classification, one of many possible, yet an intuitively plausible and still a serviceable one, was supplied on that occasion by the experimenters and involved the following four categories of meaning: connotative, denotative, associative and referential. The first consists of non- logical experiential elements, typically affective, which are evoked by words and color all other constituent elements of their meaning. The second consists of meaning elements supplied by the formal definition of the words. The third consists of whatever other experiences are evoked associatively by the encounter with a word. The last consists of the experiential elements that comprise the specific referent of the word. For example, the word "acropolis" may be understood as something that has the aesthetic properties of symmetry, simplicity and elegance or as something that entails the feeling of security that all castles entail. These experiential elements would then comprise the connotative meaning of the word. At the same time the word denotes the topographically highest point of a city, a noetic experience that those familiar with the etymology of the word would likely have. Moreover, the same word would evoke, associatively, a wide variety of other feelings and images, different ones for different people and different ones for the same person at different times. Finally, the very same word, depending on the speaker's intention and the hearer's state of mind, could refer to or evoke the notion or image of the holy precinct of Athens or of a particular ethnic diner in Manhattan, among other similarly specific notions or images. It is therefore quite obvious that if a theory claims that each meaningful experience requires its own neuronal code, given the nearly infinite range of affective colorings, of associations and of referents that a single word entails, a nearly infinite

[7] Osgood et al. 1957.

number of such codes must be postulated to inhabit each brain for just one of the tens of thousands of words that a given person may be familiar with at a given point in his life.[8] But, this issue, peripheral to the main problem under consideration, will be briefly revisited at same later point. At present, a summary of the Hebbian theory of the neuronal codes of word meaning is in order.

To begin with, there is hardly a single neuroscientist today that would contest the assumption that a different spatio-temporal pattern of neuronal signals corresponds uniquely to each experience in general and to each experience of word meaning in particular. Also, no one would contest the assumption that the pattern is constructed on the basis of a code, or of a neuronal alphabet. The notion of a unique correspondence of neuronal signal patterns and experiences is acknowledged not only by those with materialist leanings who believe that such patterns are in fact the necessary and sufficient conditions of subjective experiences of any kind, including experiences of meaning, but also by dualists who believe that such patterns may be necessary but are certainly not sufficient for experiences to arise. Universal also, is our ignorance as to the nature of the neuronal code on the basis of which cerebral mechanisms encode experiences. This is the reason the author of the article under consideration did not waist effort in dealing with the issue of the nature of the code, but focused instead on how activation patterns corresponding to each word meaning develop, how and where they are stored and how they are re-activated once established and stored.

As mentioned earlier, the proposals as to how the above processes transpire are based squarely on the well-known notions of Donald Hebb as they first made their appearance in the now classic monograph "The Organization of Behavior". In accordance, then, with Hebb's and Pulvermüller's notions which are also those that dominate in the neuroscience community, neuronal signal patterns that correspond to experiences in general and word meanings in particular, exist in two distinct states and they often transition from the one state to the other. We may name the one state passive or latent and the other active. In the latent state, the activation pattern is a virtual one and assumes the form of a circuit consisting of neurons which have acquired the capacity to discharge together, in the sense that discharging of one entails a high probability that the others in the circuit will follow suit. Activation of the neurons of the circuit, or the "cell assembly" in Hebb's terminology, transforms it into an active "reverberating circuit". The circuit in this activated state reverberates for some time; that is, it has some minimal duration which coincides with the duration of the corresponding experience or, in the case in point, with the duration of the corresponding experience of word meaning.

Subsequently, the assembly returns to its passive state and it becomes, once again, a dormant circuit until its next reawakening which will result in the re-experiencing (recall or recognition) of, approximately, the same meaning. In that

[8] The reason being that in each instant the experience is bound to be a little different than at any other instant, given the definition of "meaning" above and also given that with the passage of time, no circuit remains exactly what it was before, following the inexorable law that ordains that all living systems age and change with time.

latent state, the cell assembly can not be distinguished from others that encode and correspond to other meanings and other experiences, more generally. Today, at any rate, we do not have in our disposal methods of the requisite sensitivity for identifying each dormant circuit and for discriminating among the countless cell assemblies. In fact it may be impossible to ever develop such methods if the contention is true that a given neuron may belong to different cell assemblies because, in that event, to recognize what cells comprise an assembly would require activation of that assembly. But as an activated, reverberating circuit, the same cell assembly is potentially identifiable, with the various functional neuroimaging methods that are available today, if not in its entirety at least in outline, as a spatio-temporal activation pattern. And this possibility, as it will become evident later on, renders the theory under discussion a scientifically valid one, in that it is empirically falsifiable – a fact duly noted in the article under discussion.

To be sure, the codes of word meaning, in their aspect as dormant cell assemblies as well as reverberating circuits, are known in other contexts albeit with different names. Specifically, in their latent state, the stored codes of word meanings and of all sorts of other experiences constitute the contents of secondary memory, whether semantic or episodic. And, in their activated state, as reverberating circuits, the same circuits constitute the contents of primary memory, whether "immediate" or "working" (see e.g. Papanicolaou[9]). More precisely, that is according to this Hebbian model, the circuits in their activated state are necessary antecedents, or as the article's author put it, "the basis" of the experiences that constitute the contents of primary memory or of the stream of consciousness – to use James' celebrated phrase.[10]

Moreover, and always in accordance with the model under discussion, the neuronal signal pattern corresponding to each experience begins its career as a newly developing reverberating circuit set in motion by some (typically external) stimulus input. How, precisely, such a (typically gradual) development and establishment of new assemblies is achieved, is only specified up to the level of detail disclosed by statements such as the following, and no further.

> When the meaning of a concrete content word is acquired, the learner may be exposed to stimuli of various modalities related to the word's meaning or the learner may perform actions to which the word refers.[11]

For example, when someone is exposed for the first time to the word "coffee" he is likely also exposed to the sight and smell of a cup of coffee; or when one hears from her grandmother "Lisa! Don't run!" is at the same time experiencing the somatic and visual sensations associated with the act of running. In the first example, the visual and olfactory, and in the second example the kinesthetic stimuli, give rise to the corresponding sensory reverberating circuits, in addition to those that the verbal stimuli "coffee" or "run" do. And, if these neurons that respond to the particular acoustic and phonological features of the word comprising the "word from" are

[9] Papanicolaou 2006a.

[10] James 1890.

[11] Pulvermüller 1999, p. 260.

co-activated frequently with those responding to the visual and kinesthetic stimuli (or, to put it differently, to those reflecting the referential aspect of the meaning of the words) the assembly of the neurons corresponding to the word forms will be transformed so as to include those neurons corresponding to the referents of the words, thus creating a "higher order assembly" such that a content word

> ….may … be laid down in the cortex as an assembly including a phonological (perisylvian) and a semantic (mainly extra-perisylvian) part[12]

And,

> ….after such an assembly has been formed, the phonological signal will be efficient in igniting the entire ensemble including the semantic representation and, vice versa, the assembly may become ignited by input only to its semantic part…[13]

More analytically, following its formation, the new cell assembly will transition to its latent state where it will remain as an additional item of semantic memory until it is reactivated. Its activation may be a result of an external sensory signal: the acoustic signal comprising a spoken word, for example. Once this external sensory signal is phonologically analyzed, it will be transformed into a brain signal pattern similar to the phonological portion of the latent cell assembly and, as such, it will activate the entire higher order assembly. Once the entire assembly "ignites" and begins to reverberate, the meaning experience emerges.

Moreover,

> …..Hebbian associationist logic suggest…

(to the author) that,

> the cortical representations differ radically between words of different vocabulary types[14]

That is, the "semantic" component of each cell assembly should, according to the logic of the original formulation of the theory, be constituted of cells in different regions of the extra-perisylvian cortex for the different grammatical word classes, like the class of "content words" such as nouns, "action words", such as verbs and in different regions yet again, for the different subclasses of nouns (e.g. abstract vs. concrete nouns, nouns signifying visual objects and nouns signifying olfactory, auditory or gustatory events). Finally, on the grounds of the same logic, the location of cells encoding the phonological aspects of all words and the semantic features of function words should be in the left perisylvian region, whereas those encoding the semantic aspect of verbs should be in the motor region of the cortex and those of nouns that refer to tangible and visible objects in the posterior cortical regions.

In summary, with the exception of the process of the formation of the assemblies which is under-specified and the process whereby ignited or activated assemblies cause (represent, produce, reflect, or are the basis of) experiences which is totally unspecified in the article, the rest of the theory is experimentally accessible and

[12] Ibid., p. 260.

[13] Ibid., p. 260.

[14] Ibid., p. 260.

falsifiable through both, the new methods of functional neuroimaging, and the older ones of clinical neuropsychology and neurology that rely on the observation of the effects of focal lesions on selective word-finding and word-using deficits. It is therefore a perfectly typical or normal scientific theory even though several of its specific provisions have yet to be empirically confirmed. And, partly because it is experimentally accessible, this theory, in the form it was presented in the pages of the "Behavioral and Brain Sciences", it was sanctioned by the neuroscience community which, in this particular case, was represented by the 29 commentators of the target article. Now, given that during the intervening years since the publication of the article nothing has transpired of sufficient importance to change the opinion of the scientific community on there issues, I believe it is reasonable to consider the theory described therein as a valid and representative expression of the opinion of contemporary neuroscientists, linguists and psychologists.

Clearly though, acceptance of the main features of a theory does not imply acceptance of all its aspects. Many objections to it were voiced by the commentators of the original article and several amendments on particular points were proposed. I could add to those that care should be taken to address the difficulty, hinted at earlier, that the capacity of any normal human being, at any given point in time, to experience a virtually limitless variety of meanings on encountering a verbal stimulus, would seem to require a virtually infinite number of distinct cell assemblies. Such a difficulty should call for a modification of the notion that cell assemblies encode meanings and its substitution with the notion that, perhaps, what they encode are algorithms of production of various shades of meaning on the basis of a relatively small set of semantic features (see e.g. Papanicolaou[15]).

Similarly, acceptance of the theory in its general aspect does not imply that the theory is considered complete. Among the several suggestions, on the part of the commentators, of pending issues to be addressed, was the one that elicited the author's doubt that it could ever be profitably addressed, quoted at the head of this exposition: Questions such as "what renders a cell assembly a concept?" the author had cautioned, are questions that do not admit of reasonable answers. When scientists venture to address them, as many prominent ones have done, they invariably come up with unsatisfactory answers, in the sense that these answers never settle the issue. Specifically, they do not specify the efficient cause of the generation of meaning but only point to some of the, possibly, necessary conditions of the process, like particular voltage levels or spectral properties that neuronal signals must have in order for them to generate experiences or in order for some (unspecified) neuronal mechanism to turn them into experiences. Yet, the same author, compelled by the same justifiable urge as the one motivating those he warned, could not resist the temptation to also provide an answer: The physical basis of conscious meaning, he said, or the necessary conditions for the production of conscious experience are (a) cortical cell assemblies, (as opposed, I suppose, to subcortical?); (b) assemblies involving many neurons; (c) assemblies involving neurons possibly linked strongly and (d) triggers to ignite the assemblies.

[15] Papanicolaou 2006b.

This list of conditions, whether one may consider it more or less complete, or more or less accurate than similar lists compiled by other authors remains, nevertheless, merely a list of antecedent conditions that does not address the essential aspect of the issue: the specification of the efficient cause, that is, the neuronal mechanism that transforms signals into meanings or the process whereby assemblies with those properties cause experiences directly, by themselves. A list moreover, that, much like the other lists, raises further questions like: why should strongly connected assemblies

> …make me experience consciously?[16]

Why, indeed.

As I stated at the outset, I believe that what renders the issue enigmatic and what renders such questions as the ones the author mentioned, confusing, is that the way the issue is articulated precludes all answers except the one of a causal relation between brain signals and experiences; the one, moreover, mandated not by any empirical evidence but by a metaphysical belief, the repercussions of which we end up finding unconvincing and confusing. And, as I also noted, the enigma can be resolved. But to resolve it we have first to step back from the reflexive, easy solutions that its form of statement automatically suggests to us.

In the article under discussion, the way the issue is stated leads us to only consider either of two causal relations between brain signals and meanings: the signals are necessary or the signals are sufficient to cause meaning. The author does not care to distinguish them, possibly because he considers them both equally confusing and unsatisfactory. Both of them however, are implied in his conclusion that the brain signals constitute the basis of the experiences of meaning, which creates two riddles: the one is that of the emergence of meaning out of the patterns of brain signals; the other is the riddle of how some brain mechanism or other transforms these signal patterns into experiences. In either case, the first step towards resolving the enigma is to realize the logical impossibility that neuronal codes can produce experiences either by themselves or through the contribution of other neuronal networks. And there is no better way of accomplishing this sort of feat than through the fabulous invention of Aesop: the fable.

The fable I am about to relate as the most relevant to the enigma in question is the one that has been told with the incomparable elegance and lucidity of Gabriel Garcia Márquez in the "One Hundred Years of Solitude".[17] Therefore, I have two reasons to relate it at length: the one is its convincing power and the other the sheer pleasure of recounting it. It is the story of the plague of insomnia, the amnesia for the meaning of words (among other experiences) that followed it, and the ingenious, but none the less futile attempt of the heroes of the story to rescue the meaning of words from oblivion by means of signs and signals:

> It was Aureliano who conceived the formula that was to protect them against loss of memory for several months. He discovered it by chance. An expert insomniac, having been one

[16] Pulvermüller 1999, p. 326.

[17] Márquez 1967/2006.

> of the first, he had learned the art of silverwork to perfection. One day he was looking for the small anvil that he used for laminating metals and he could not remember its name. His father told him: "Stake." Aureliano wrote the name on a piece of paper that he pasted to the base of the small anvil: stake. In that way he was sure of not forgetting it in the future. It did not occur to him that this was the first manifestation of a loss of memory, because the object had a difficult name to remember. But a few days later he discovered that he had trouble remembering almost every object in the laboratory. Then he marked them with their respective names so that all he had to do was read the inscription in order to identify them. When his father told him about his alarm at having forgotten even the most impressive happenings of his childhood, Aureliano explained his method to him, and José Arcadio Buendía put it into practice all through the house and later on imposed it on the whole village. With an inked brush he marked everything with its name: *table, chair, clock, door, wall, bed, pan*. He went to the corral and marked the animals and plants: *cow, goat, pig, hen, cassava, caladium, bananas*. Little by little, studying the infinite possibilities of a loss of memory, he realized that the day might come when things would be recognized by their inscriptions but that no one would remember their use. Then he was more explicit. The sign that he hung on the neck of the cow was an exemplary proof of the way in which the inhabitants of Macondo were prepared to fight against loss of memory: *This is the cow. She must be milked every morning so that she will produce milk, and the milk must be boiled in order to be mixed with coffee to make coffee and milk*. Thus they went on living in a reality that was slipping away, momentarily captured by words, but which would escape irremediably when they forgot the values of the written letters.[18]

when, that is, they would forget the meaning of the signs, or the meaning of each element of the code that constituted the names and the definitions of things. This is, then, the fable. I believe it makes it obvious that signs and signals, however elaborate, will never suffice to create meaning in the absence of the decipherer, the bearer of the meanings that the signs remind him; the efficient cause of the conversion of signs into meanings; the agent who knows the code, and who can read and interpret it.

If Márquez's fable did not succeed in persuading the undecided of the insufficiency of the signals or the reverberating circuits alone, to give rise to meaning and of the necessity of an entity that can function as interpreter of the signs, any additional attempts of my own are certainly bound to fail. Consequently, in order to continue, I can only assume that it is clear to all that the list of properties that brain signals should have in order to be interpreted, no matter if it were exhaustive, would not suffice to explain the emergence of experiences. To explain such emergence, specification of their properties must be supplemented by the specification of the efficient cause, in this case the specification of the entity that interprets them, whether they are patterns of electrochemical processes in brains, engravings on stones, or markings, in the common alphabet, pasted on anvils, cows, pigs, hens, or bananas.

But if the only entities discernible within brains are formations of electrochemical signals, what would be the entity that fulfills the role of that mysterious efficient cause? – "The mind, of course", is the answer of the dualists from the time of Descartes onward, since it is only minds, as far as we can ascertain, that interpret codes and signals of any sort. That answer resolves the riddle in a way that evidently

[18] Ibid., pp. 46–47.

satisfies a considerable portion of humanity and even some neuroscientists. Ultimately, it may even turn out to be the correct one. Yet I must ignore it here, for two reasons: First, because, since Descartes' failed attempt to explain how the mind interacts causally with the brain, there has been no other proposal to that effect that has proved to be any more credible, and unless such a proposal is made, the way minds read brain signals will remain as much of a mystery as the way minds cause brain activation that, in turn, causes overt action. Second, I ought to ignore this resolution of the enigma, simply because I am unable to conjure up anything that improves on Descartes' proposal.

I trust it is clear that such reasons as the above for rejecting this explanation belong to the type routinely used in the natural science, if not in metaphysics also, for rejecting particular explanations of empirical data. So, the question remains: what in the brain could serve as a substitute for mind? Those that have ventured to guess have proposed the "executive" functions, the brain mechanisms of which are localized, as far as we can tell, in the frontal lobes and which, other colleagues with a flair for the opt metaphor have named "General"[19] or "Interpreter".[20] But whether the mind surrogate is called "executive algorithm" or more poetically "General" or "Interpreter", it cannot furnish a satisfactory answer to the enigma for a very simple reason: Any and all brain mechanisms, no matter how sophisticated, produce one and only one type of output: signals and signal patterns; never meanings, as far as anyone has been able to witness and record.

Therefore, we are faced with the all too familiar scenario of the infinite series of "Generals" or "Interpreters" one embedded into the other, where each reads and interprets the signals that the previous one produces, in the off-chance that one of them may produce meanings instead of signals. And this is where the issue stands: In spite of the tremendous progress in the neurosciences and in spite of the rise of neurophilosophy, the discipline which is exclusively invented for settling it on the basis of the evidence furnished by neuroscience, the world remains apparently unconvinced since it continues to be as divided on this issue as it has always been. Some people persist in the belief that brain signals suffice to generate conscious experiences, especially if they are the output of "executive" mechanisms, but the rest continue to uphold the notion that to become sufficient in engendering experiences, brain signals require the added contribution of a "mind". Only, as I have already mentioned, it has proved impossible for anyone to suggest the process whereby a presumably non-material force, "mind", "spirit" or whatever else one might care to call what James[21] named "fiat of the will", ignites the brain cells and throws the motor mechanisms of the brain into action.

But, of course, neither has any of the dozens of "neuroscientific" proposals (see e.g. Chalmers[22] for a partial list) offered any sufficiently persuasive model of the kind of causal process that turns patterns of electrochemical signals in the brain into

---

[19] Gazzaniga 1992.

[20] Ramachandran and Blakeslee 1998.

[21] James 1888/1983.

[22] Chalmers 2000.

experiences. Certainly many ingenious verbal devices have been fashioned to soften the shock of incredulity one is liable to experience any time one reflects on statements to the effect that such and such brain signals generate meaning. But the most ambitious of them all appears to be the now nearly ubiquitous notion of "emergence": The way water with its properties of wetness, etc. emerges from atoms of hydrogen and oxygen that do not have these properties, is also the way thoughts and sensations, with all their subjective properties not present in their antecedent brain events, emerge from them, somehow.

The difference is, though, that there is not shock of incredulity when one contemplates how a perfectly material thing or substance, like water, emerges out of also perfectly material antecedents. The shock is only felt when thoughts are said to depolarize brain cells or when ion movements in those cells are said to create meanings. And it is felt in these cases exclusively, for two reasons: First, because the relation of brain events and experiences offers the only example where the emergent entity is by definition and by nearly universal consensus radically different from the entities out of which it emerges. Second, because the notion of "emergence" as it applies to all cases is neither a clear nor a simple one. The way it occurs in the literature, it entails two intertwined concepts referring to two very different things. The first concept refers to the phenomenon itself whereby an entity arises from antecedent phenomena that lack some of the properties that characterize it. The second concept is meant to point to the process that transforms the latter into the former by insinuating that this process is not the ordinary one of causation but a different one, albeit different in unspecified ways. And, both concepts fused together are meant to bridge the otherwise unbridgeable gap of incredulity that the frank expression "brain signals cause meaning" creates, as the following quotation from an early proponent of emergence, Roger Sperry[23] amply illustrates:

> …. the subjective mental phenomena are conceived to influence and to govern the flow of nerve impulse traffic by virtue of their encompassing emergent properties…. The individual nerve impulses and associated elemental excitatory events are obliged to operate within larger circuit-system configurations. These…. Interact causally with one another at their own level as entities. It is the emergent dynamic properties of certain of these higher specialized cerebral processes that are interpreted to be the substance of consciousness.[24]

The first constituent concept of the notion of "emergence", in so far as it refers to the fact that the emergent entity has properties not found in its antecedents, cannot be contested. It can and should be contested, however, in so far as it is transmuted into its twin; in so far as it is meant to imply that the process is other than causal, because nowhere has anyone ever specified in exactly what the difference between the process of causation and the process of emergence consists and how is that difference to be empirically ascertained. Therefore, since there is no specified difference between causation and emergence to be detected, these two concepts should be, as far as empirical science goes, identical.

[23] Sperry 1969.

[24] Ibid., p. 534.

At this point one cannot refrain from commenting on the obvious, even at the risk of diverting the course of this argument. And it is patently obvious that neither ordinary causation nor emergence may ever bridge the divide to the satisfaction of most reasonable individuals, as long as the purely metaphysical notion that brain events (or trees and stars, for that matter) belong to one ontological category, whereas experiences like thoughts and sensations, but also experiences or percepts like brain activation patterns, trees and stars, belong to another altogether different category; one defined by the absence of the defining features of the former! But they do not. This, at any rate, is the realization at the core of the promised solution to the riddle: A real stone, a real tree, a real brain activation pattern that we see and touch are just that, percepts, that is, empirically determined phenomena. What exactly is behind them that, interacting with our sensory apparatus (or our sensory apparatus-enhancing instruments) gives rise to them, is a theoretical matter. And, whereas theories about the ultimate nature of trees, stars and stones change, real stores, trees and stars remain as ever, perceptual phenomena.

As for the solution of the enigma, it is this: The pattern of brain signals, or the reverberating circuits, the spatio-temporal outline of which is potentially visualizable with the modern methods of functional neuroimaging, is one of two aspects of the same reality, the second aspect being the experience of meaning. The solution is, in other words, that signals and experiences have a relation other that the causal one that the dominant neurophilosophical dogma posits to be the only one admissible. But the proposed solution, the kernel of which could be found in the thought of Spinoza but of Kant as well is, on the one hand, so diaphanous and on the other hand so much obscured by metaphysical discussion since Spinoza's time, as to have escaped notice of most neuroscientists. I will try to argue, though, that of all proposed relations between brain events and experiences, is the one most consistent with pragmatic thinking and with the results of the application of the methods of natural sciences.

There is only one logical prerequisite, external to my argument, for accepting this solution. And that prerequisite is adherence to the assumption, already mentioned, which conflicts neither with the materialist or the dualist variety of metaphysical dogma and is one of the corner-stones of functional neuroimaging: the assumption that to each individual experience in general and each experience of meaning in particular corresponds a unique brain activation pattern (or activated cell assembly or reverberating circuit). But, because empirical verification of the assumption is not at hand, my argument must take the form of a thought experiment:

I am a participant in a neuroimaging experiment involving an instrument that records with high temporal and spatial fidelity the patterns of my brain's activation contingent on my hearing words. The experimental set-up also allows me to view, if I so choose, those spatio-temporal activation patterns as they develop in real time and as they appear in the screen of the imaging device. The aforementioned assumption being valid, I have two experiences at any given time: the visual experience of a spatio-temporal pattern (representing, let us say, the configuration of moving ions throughout my brain) and a complex temporal experience consisting of auditory sensations representing the heard "word form" along with its meaning.

Having had several such pairs of experiences, I realize that there is indeed a remarkable concordance between them: Each word heard and understood is associated with a different spatio-temporal pattern of activation that appears to be unique to it. To ascertain that the activation patterns in my brain do not correspond only to the sounds of the words and they are not, for that reason, related only to the word forms, and to discover whether there is concordance of activation patterns and the experienced meaning as well, I design and subject myself to a series of further experiments. In them, the same words are sometimes heard and at other times read off a screen. Moreover, after each presentation of each word, I recount and record, in as great detail as I reliably can, the different shades of meaning that each of the several presentations of the same word evoked in me.

This time I observe that the early part of the activation patterns are constant across the different visual (and the different auditory) presentations of the same word but are still different and unique to each word. Moreover, these early potions of the activation patterns would allow any outside observer to classify them in two categories: those that were evoked by the auditorily presented words in one category and those evoked by the visually presented ones in another, on the basis of a gross difference in their spatial aspect, the former featuring higher levels of temporal lobe activation and the latter greater levels of posterior cortical activation (see e.g. Papanicolaou et al.[25]).

As for the later parts of the activation patterns, I find that they again differ reliably between words, being specific to each, regardless of whether the words were visually or auditorily presented, indicating that the later part of the activation pattern co-varies with the word meaning and not the word form. So that, conceivably, an external observer could tell, just by looking at the latter part of the activation pattern, what word was understood and not merely heard. I also find that there are further differences among the later part of the activation patterns associated with each presentation of any given word. Yet these are intractable being never repeatable, following, in that respect, the shading of the experienced meaning that was also unique and non-repeatable from one presentation of the same word to the next[26] (see e.g. Papanicolaou,[27,28] for an explanation).

Also intractable was the point in time dividing the early part of the activation pattern, specific to the word form, from the subsequent, meaning-specific part of the pattern. This is so and must remain so even in thought experiments, if the latter are to be meaningful, due to the so-called "prior entry effect" that has been recognized since the inception of experimental psychology.[29] This effect, demonstrated repeatedly in the context of dozens of experiments over more than a century and a

[25] Papanicolaou et al. 2009.

[26] As commented in note 8 above, given that each meaning is uniquely associated with a circuit and that circuits, being physical entities, age or otherwise change with time, no circuit, therefore no experience of meaning can be absolutely the same at two successive time points.

[27] Papanicolaou 1998.

[28] Papanicolaou 2007.

[29] Wundt 1874.

half, consist in the fact that when one is to time the relative onset of two (or more) temporally adjacent or overlapping events, one deems the event one happens to attend more, as the earlier one. And, in situations as those of the thought experiment just recounted, attention invariably shifts unpredictably, each time a word is presented, from the experience of its meaning to the percept of the corresponding activation pattern, making it impossible to tell at what point in the temporal unfolding of the pattern does one cease to be aware of the word form and becomes aware of the word meaning. The experiment, therefore, in the event it was in fact conducted, would only demonstrate that there is an extremely high correlation between two categories of quasi-simultaneous phenomena of comparable duration: Brain activation patterns (or reverberating circuits) and word meaning.

And the question now becomes, how are we to interpret these empirical observations that resulted (or would result) from the application of the methods of the natural science without staying into metaphysics? I remind the reader that the correlation of the two phenomena may not be interpreted as causal, whereby the brain signals cause meanings or in the sense that the meanings "emerge" out of the signals, for reasons that were detailed earlier. The same reasons would militate against such a causal explanation even if it were possible to empirically determine that the meaning-specific part of the activation pattern preceded the onset of the experience of meaning[30] in the way Libet[31] thought he was able to determine the relative onset of two experiences in his often- discussed experiment.[32]

But if causality must be excluded, what other explanations of the relation of the two classes of phenomena are possible? One explanation could be that of psychophysical parallelism (see e.g. Boring's "History"[33]). This explanation is certainly compatible with the experimental facts and is, perhaps, a reasonable one. A second one is the explanation based on that variant of the theory of "neural identity" that interprets the covariance of the two types of phenomena (meaning and activation pattern) as an indication of their identity qua phenomena. That theory, incidentally, was extensively discussed in the 1950s (see e.g. Feigl[34]) it continues to be debated, and some of its variants may be said to coincide with the third explanation of the relation, the one I favor and the one attributed to Spinoza, stating that the two distinct phenomena are correlated because they are different aspects of the same underlying process.

We must now select the most reasonable explanation, the one that accords best with the spirit and practice of empirical inquiry. I would, accordingly, reject the first explanation: In no other circumstances has the co-variation of a virtually infinite series of pairs of phenomena been attributed to mere coincidence. I reject with even

[30] Although (efficient) causes must precede their effects, mere temporal precedence alone does not qualify an event as a cause of another subsequent to it.

[31] Libet 1985.

[32] That Libet did not, in fact, accomplish that deed see more recent commentary by Pockett 2002, and Breitmeyer 2002.

[33] Boring 1950.

[34] Feigl 1958.

less reservation the "identity" explanation because it is at odds with the empirical facts: We are witnessing two distinct classes of phenomena. The one class consist of visual percepts; the other of meanings. To say that in spite of the fact that, empirically, we are in the presence of two phenomena we are nevertheless in the presence of one is, at best, to allow a metaphysical belief distort beyond recognition the empirically ascertained facts.

There remains the third explanation. By adopting it, we admit the obvious: We admit the presence of two distinct but highly, if not perfectly, correlated sets of phenomena. And, whenever empirical observation discloses highly correlated series of phenomena, such as the phenomena of thunder and lighting, it has been normal practice in science to refer the two to a common, third process. Thus we refer the correlated phenomena of lighting and the thunder to electrical processes in the clouds, although in other cases, like in the case under discussion, the common underlying reality, unlike electricity, may be either temporarily unknown or permanently, that is, in principle, unknowable.

Thus we have arrived at the resolution of the enigma. Some, however, may remain skeptical and unsettled because the common reality behind meanings and brain signal patterns appears to be in principle occult. But should the neuroscientist, the linguist or the psychologist, in their role as natural scientists, be so concerned with the fact that some realities are in principle unknowable as Immanuel Kant[35] had argued and as modern natural philosophy is reasserting (see e.g. Eddington[36] and Russell,[37,38]) as to reject otherwise perfectly reasonable interpretations of observed facts? I believe not. Not any more, at any rate, than the physicist should be concerned with the experimentally unapproachable, and in principle unknowable, nature of the reality, expressions of which are the "particle" and the "wave". And, the same way that the physicists managed, after more than half a century's study and debate, to internalize how unreasonable it would be to consider the one phenomenon as the cause of the other, so should the behavioral scientists finally realize how unproductive their models are bound to remain when they entail the demand that reverberating circuits must be shown to be the causes of word meanings or of experiences at large. But once this absurd demand and the need of an agent, or a decipherer of signal patterns, is eliminated, the neuroscientists can and should _ redirect and refocus their efforts to the eventual discovery of what cell assemblies, how numerous and how strongly interconnected and how widely distributed, communicating with signals of what intensity, and reverberating at what frequencies are the one aspect of the occult reality of which the other aspect are experiences, since neither can be shown empirically to "cause" or "emerge from" the other.

One may protest though, that this is an unnecessary exhortation since scientists of whatever metaphysical leanings do precisely that anyway: they strive to identify the necessary conditions of phenomena that interest them. This is true but it is also true

[35] Kant 1781, 1787/1996.

[36] Eddington 1939.

[37] Russell 1914.

[38] Russell 1948.

that many of them also practice at the same time metaphysics and do so unwittingly. There is, of course, nothing untoward in vaulting beyond the region illuminated by the empirical evidence and inquiring, for example, about such things as whether the hidden reality is more like ordinary matter (as many, if not most, believe) more like mind (as hardly anyone believes) or like neither (as the "neutral monists", like Russell,[39] believe). On the contrary, reaching beyond the evidence in an effort to discern their deeper significance, besides being a natural urge, is a highly prized and commendable exercise. But it is such only as long as it is performed competently and with full awareness of the fact that it is a metaphysical one.

References

Boring EG (1950) A history of experimental psychology. Appleton-Century-Crofts, Inc., New York
Breitmeyer BG (2002) In support of Pockett's critique of Libet's studies of the time course of consciousness. Conscious Cogn 11:280–283
Chalmers DJ (2000) What is a neural correlate of consciousness? In: Metzinger T (ed) Neural correlates of consciousness empirical and conceptual questions. The MIT Press, Cambridge, MA
Churchland PS (1986) Neurophilosophy: toward a unified science of the mind-brain. The MIT Press, Cambridge, MA
Churchland PS (2002a) Brain-wise: studies in neurophilosophy. The MIT Press, Cambridge, MA
Churchland PS (2002b) Braintrust: what neuroscience tells us about morality. Princeton University Press, Princeton
Crick F, Koch C (1990) Towards a neurological theory of consciousness. Semin Neurosci 2:263–275
Eddington AS (1939) Philosophy of physical science. Cambridge University Press/The MIT Press, Cambridge/Cambridge, MA
Feigl H (1958) The 'mental' and the 'physical'. In: Feigl H, Scriven M, Maxwell G (eds) Concepts, theories and the mind-body problem, Studies in the philosophy of science. University of Minnesota Press, Minneapolis
Gazzaniga MS (1992) Nature's mind, the biological roots of thinking, emotions, sexuality, language, and intelligence. Basic Books, New York
Hebb DO (1949) The organization of behavior. Wiley, New York
James W (1888/1983) What the will effects. In: Burkhandt F, Bowers F (eds) Essays in psychology. Harvard University Press, Cambridge, MA
James W (1890) The principles of psychology. Dover Publications, New York
Kant I (1781;1787/1996) The critique of pure reason. Hackett Publishing Company, Inc., Cambridge, UK
Koch C, Crick F (1994) Some further ideas regarding the neuronal basis of awareness. In: Koch C, Davis J (eds) Large-scale neuronal theories of the brain. The MIT Press, Cambridge, MA
Libet B (1985) Unconscious cerebral initiative and the role of conscious will. Behav Brain Sci 8:529–566
Márquez GG (1967/2006) One hundred years of solitude. Harper Perennial Modern Classics, New York
Osgood CE, Suci GJ, Tannenbaum PH (1957) The measurement of meaning. University of Illinois Press, Urbana
Papanicolaou AC (1998) Fundamentals of functional brain imaging. Swets & Zeitlinger, London

[39] Ibid.

Papanicolaou AC (2006a) Chapter One: Phenomena and constructs. In: Papanicolaou AC et al (eds) The Amnesias: a clinical textbook of memory disorders. Oxford University Press, Oxford/New York

Papanicolaou AC (2006b) Chapter twelve: Notes for a theory of memory. In: Papanicolaou AC et al (eds) The Amnesias: a clinical textbook of memory disorders. Oxford University Press, Oxford/New York

Papanicolaou AC (2007) What aspects of experience can functional neuroimaging be expected to reveal? Int J Psychophysiol 64:101–105

Papanicolaou AC et al (2009) Clinical magnetoencephalography and magnetic source imaging. Cambridge University Press, Cambridge

Pockett S (2002) On subjective back-referral and how long it takes to become conscious of a stimulus: a reinterpretation of Libet's data. Conscious Cogn 11:144–161

Pockett S, Banks WP, Gallagher S (eds) (2009) Does consciousness cause behavior? The MIT Press, Cambridge, MA

Pulvermüller F (1999) Words in the brain's language. Behav Brain Sci 22:253–336

Ramachandran VS, Blakeslee S (1998) Phantoms in the brain: probing the mysteries of the human mind. William Morrow, New York

Russell B (1914) Our knowledge of the external world as a field for scientific method in philosophy. Open Publishing, Chicago/London

Russell B (1948) Human knowledge: its scope and limits. George Allen & Unwin, London

Sperry RW (1969) A modified concept of consciousness. Psychol Rev 76:532–536

Wundt W (1874) Grundzüge der physiologischen psychologie. Wilhelm Engelmann, Leipzig

Chapter 13
Alfred North Whitehead and the History of Consciousness

Laura Hyatt Edwards

> *There remains the final reflection, how shallow, puny, and imperfect are efforts to sound the depths in the nature of things. In philosophical discussion, the merest hint of dogmatic certainty as to finality of statement is an exhibition of folly.*
>
> Whitehead (1929a, p. xiv)

13.1 Placing Whitehead in Historical Context

Alfred North Whitehead's natural philosophy is a notable addition to the history of Chalmers' hard problem of conscious experience[1] because Whitehead rejected entirely the material basis of Cartesian dualism and replaced it with an event ontology Epperson recently described as similar to twenty-first century 'consistent histories' quantum mechanics theories.[2] Although Whitehead did not set out specifically to tackle the hard problem he saw such problems as a test of the coherence and adequacy of any natural philosophy. One reason his work may be difficult to grasp is that he unpacked scientific assumptions so long-held it is difficult, even today, to see them as anything but brute fact. A portion of the secondary literature Whitehead's work engendered concerns the role of quantum mechanics in consciousness.[3] Some argue that his ontology was misunderstood, however, suggesting that its implications for the neuroscience of conscious experience have not been fully explored.[4]

[1] Chalmers 1995, p. 201.

[2] Epperson 2004, p. 1. See also Epperson 2009.

[3] For example see Penrose 2011; Petkov 2010, Pred 2005; Stapp 2005; Velmans 2009; Weber and Weekes 2009.

[4] Epperson, *op cit*.

L.H. Edwards (✉)
Department of Psychology, East Carolina University, Greenville, NC 27858, USA
e-mail: edwardsla@ecu.edu

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_13,

At 23 years of age, Whitehead's 1884 dissertation on Maxwell's equations for electricity and magnetism at Trinity College won him a fellowship and lecturer position there.[5] Subsequently, he published two works on viscous fluids in 1889.[6] In September of 1905, the same month and year Einstein submitted his "$e = mc^2$" paper,[7] Whitehead submitted a memoir on mathematical concepts of material in space to the Royal Society of London.[8] The introduction of this memoir described it as, "concerned with the possible relations to space of the ultimate entities which (in ordinary language) constitute the "stuff" in space."[9] In 1912, when Whitehead was 51, he submitted a job letter to the University College of London seeking Karl Pearson's recently vacated position. In this letter he described his 22 year career by saying it, "has always had as its ultimate aim the general scrutiny of the relations of matter and space and the criticism of the various applications of mathematical thought."[10]

Although the majority of his published work stemmed from this theme, until his letter emerged in 1975, few scholars realized that Whitehead's interest in physics had shaped much of his scholarship.[11] Many still think of Whitehead as co-author of *Principia Mathematica* with Bertrand Russell while he was at Trinity.[12] Because he completed *Science and the Modern World* and *Process and Reality* while he was a Harvard philosophy professor, others associate him primarily with his philosophical writing.[13] He was only at Harvard for the end of his career, however, from 1924 until he retired in 1937.[14] It may seem incongruent with his reputation as a mathematician and philosopher to assert that by 'scientific object' and process in *Process and Reality* Whitehead was referring to a process of quantum mechanical state evolution.[15] Yet his wish to apply mathematical understanding to the physical world dominated his career and was entirely consistent with the idea that he saw ontological relevance in Schrödinger's equation.

Whitehead's place in the history of consciousness is noteworthy in part because it was never his overarching goal to explain consciousness and in part because his thinking moved against materialist currents of thought among his contemporaries. His was a speculative philosophy of natural science, not an epistemology.[16] As physicists such as Einstein, Heisenberg, and Schrödinger developed their ideas about space, time, material and quantum mechanical events,[17] Whitehead's work led him

[5] Lowe 1985, p. 10.

[6] Whitehead 1889a, b.

[7] Einstein 1905b.

[8] Whitehead 1906.

[9] *Ibid*, p. 1.

[10] Lowe 1975, p. 86.

[11] *Ibid,* pp. 87–88.

[12] Whitehead and Russell 1910–1913.

[13] Whitehead 1925, 1929a, b.

[14] Lowe 1985, p. 133.

[15] Epperson 2004, p. 1.

[16] Whitehead 1929a, b, p. 18.

[17] Einstein 1905a, b; Heisenberg 1927; Schrödinger 1926.

to doubt several fundamental scientific assumptions. He questioned, among other things, an unacknowledged reliance on empty receptacle space[18] and the logic of assuming disconnected material could exist in space[19] or that one can locate and precisely measure material in space with immaterial geometric points.[20] Further, he disagreed with the reigning philosophy of science in his era, logical positivism, warning that a view that science can only concern itself with directly observable entities would "… produce a timid, shut in, unenterprising state of mind, engaged in the elaboration of details."[21]

Consistent with the conflict between his own and contemporary scientific thought, he also disagreed with the trend among U.S. psychologists to dismiss mind, challenging a young Burrhus Frederick Skinner to, "… describe in behavioral terms a green dragon on the dining-room table that wasn't there."[22] While neuroscientists such as Broadman, Cajal, Lashley, and Sherrington examined mechanisms of neural communication, structure-function relationships, and reflexes Whitehead repudiated the materialist basis of the mind-body problem.[23]

In *Process and Reality* Whitehead described his wish to develop a speculative philosophy that, "… is the endeavour to frame a coherent, logical, necessary system of general ideas in terms of which every element of our experience can be interpreted."[24] To him this meant that, "… everything of which we are conscious, as enjoyed, perceived, willed or thought, shall have the character of a particular instance of the general scheme."[25] By coherent he meant that its fundamental ideas ought to presuppose each other, be meaningless without each other, and therefore be mutually implicative rather than mutually exclusive.[26] He sought to divest his thinking of all assumptions until only the fewest necessary ideas remained.

Whitehead's pursuit of coherence meant he opposed acquiescence to explanatory gaps between levels of analysis. Such gaps occurred because scientists accepted as brute fact contradictory first principles such as the idea that animated entities could be made of fundamentally inert, disconnected, material. To Whitehead the intractability of the mind-body problem arose from a long history of accepting incoherence in natural philosophy and therefore represented a test of his ideas.[27] Because Whitehead grounded his philosophy in his ontological framework, a brief introduction to his ideas may clarify the subsequent summary of his view of consciousness.

---

[18] Marvin 1927–1928, p. 153. See also Whitehead 1929a, b.

[19] Whitehead 1934, pp. 32–37.

[20] Whitehead 1906, p. 1.

[21] Whitehead 1936, in a recently discovered letter to his former student, Henry Leonard.

[22] Whitehead 1963, pp. 99–100. See also Snyder 1990 for Skinner's memory of this encounter.

[23] Broadman 1909/1999; Cajal 1906; Lashley 1924; Sherrington 1925, Whitehead 1920/1955, Chapter 2, "Theories of the Bifurcation of Nature" and, for example, p. 185.

[24] Whitehead 1929a, b, p. 18.

[25] *Ibid*.

[26] Epperson 2009, p. 340.

[27] Whitehead 1920/1955, p. 27.

13.2 Brief Introduction to Whitehead's Natural Philosophy

Scholars have described Whitehead as a metaphysicist, an epistemologist, or an onto-epistemologist but the multiple meanings ascribed to philosophical terms may have obscured his stated purpose.[28] Whitehead thought one source of philosophical difficulty was that earlier philosophers had unwittingly tried to develop an epistemological and an ontological framework at the same time so his starting point was admittedly an epistemological stance. This was only because he had to start somewhere. He wanted to make the fewest assumptions possible and it was for this reason that he chose to take for granted only that sense data indicated a vague something going on. Whitehead anticipated the misunderstanding that he was developing a metaphysical stance regarding mind outside of or in addition to nature because he separated metaphysics from his work on natural philosophy this way,

> The immediate thesis for discussion is that any metaphysical interpretation is an illegitimate importation into the philosophy of natural science. By a metaphysical interpretation I mean any discussion of the how (beyond nature) and of the why (beyond nature) of thought and sense-awareness. … It is the philosophy of the thing perceived, and it should not be confused with the metaphysics of reality of which the scope embraces both perceiver and perceived. No perplexity concerning the object of knowledge can be solved by saying that there is a mind knowing it.[29]

It was from explicit recognition of one's natural philosophy that one could begin to speculate how planetary motion, living things, a sunset, or qualia fit into nature.[30] Scientists and minds are simultaneously 'in nature' along with everything else and it was wise to remember that no one observes nature from the outside looking in. He concluded that it did not matter what topic a scientist wanted to study, the essential problem is the same—how to show its relatedness to the rest of nature.

To Whitehead, a natural philosophy was hiding behind the skirts of metaphysical argument when it excluded from science anything of which one was aware, designated phenomena 'unknowable' because of mind, or blamed mind for explanatory difficulties.[31] When physicists were debating the puzzling behavior of particles in the early days of quantum mechanical experimentation most suggested epistemological interpretations implying that scientists would never be able to know objective reality for various reasons.[32] Whitehead addressed the idea among physicists that conscious minds were necessary to determining the outcomes of quantum mechanical state evolution saying, "It comes to this, that there has been so much happening, and that, so far as we know, there is not enough mind to go round."[33] Later he asserted,

---

[28] Weber 2006, p. 117.

[29] Whitehead 1920/1955, p. 28.

[30] *Ibid*, p. 29.

[31] *Ibid*, p. 46.

[32] Epperson 2004, pp. 25–31.

[33] Carr et al. 1922, p. 132.

> My argument is that this dragging in of the mind as making additions of its own to the thing posited for knowledge by sense awareness is merely a way of shirking the problem of natural philosophy. That problem is to discuss the relations *inter se* of things known, abstracted from the bare fact that they are known. Natural philosophy should never ask what is in the mind and what is in nature. … It may be that the task is too hard for us, that the relations are too complex or too various for our apprehension … But at least let us not conceal failure under a theory of the byplay of the perceiving mind.[34]

Accordingly, as new findings came to light in quantum mechanics, Whitehead incorporated their implications. This led him to speculate that scientific objects (e.g., electrons, particles, waves) are events and not discrete objects at all. Whitehead asserted that, "A classification of natural entities is the beginning of natural philosophy."[35] Bare sense awareness of 'something going on' yields two things to consider, the discerned and the discernible. The discerned are entities one discriminates with individual peculiarities and the discernible are the vague and poorly defined background of relata. "This character may be metaphorically described by the statement that nature as perceived always has a ragged edge…"[36] At that time scientists were more likely to see ragged edges as the outcome of faulty measurement to be trimmed away with increasingly precise instrumentation. Many saw the following as brute fact: that bits of inert material (matter, electrons, electricity) existed disconnected from other bits of material, that empty space existed, and that material could move through this empty space.[37]

Passages of *Science and the Modern World* and *Process and Reality* read like textbooks on the history of science because Whitehead was acutely aware that alternative assumptions about nature swing like a slow pendulum over this history's head.[38] Whatever scientists assumed about space, time, and material pushed aside important questions, leaving them unanswered if not unasked. What is space, if anything? Gravity? How does anything exist without being *in* space? How do things change location if space is empty?

> The history of philosophy discloses two cosmologies which at different periods have dominated European thought, Plato's *Timaeus*, and the cosmology of the seventeenth century whose chief authors were Galileo, Descartes, Newton, and Locke. … it is wise to follow the clue that perhaps the true solution consists in a fusion of the two previous schemes, with modifications demanded by self-consistency and the advance of knowledge.[39]

Whitehead did not claim to have invented the philosophy of organism but argued it was founded on philosophical traditions of Plato and Aristotle, Descartes, Hume, and Locke. His philosophy of organism, however, was "just those elements in the

---

[34] Whitehead 1920/1955, p. 30.

[35] *Ibid*, p. 49.

[36] *Ibid*, p. 50.

[37] *Ibid*, p. 26.

[38] Whitehead (1925, p. ix) described *Science in the Modern World* as "a study of some aspects of Western culture in the past three centuries so far as it has been influenced by the development of science."

[39] Whitehead 1929a, b, p. xiv.

writings of these masters which subsequent systematizers have put aside."[40] The balance Whitehead sought is represented in several forms throughout *Process and Reality* including substance and flux,[41] divisible but undivided, the real potential expressed in continuity and the actual.[42] Plato's *Timaeus* captured many of the ideas of nature Whitehead wanted to retain. Platonic container space is not readily distinguishable from material and had a generative relationship to the rest of nature unlike the empty receptacle space retained in Newtonian mechanics. Platonic space was a container or *choros*. Its Greek origins are not that of a vacuum or empty thing in which to place objects from somewhere else, however, but that of a body cavity with connotations of a womb.[43]

> In our former discussion, I distinguished two kinds of being … But now a third is required, which I shall call the receptacle or nurse of generation … The containing principle may be likened to a mother, … space or matter is neither earth nor fire nor air nor water but an invisible and formless being which receives all things…There is also a third nature—that of space, which is indestructible, and is perceived by a kind of spurious reason without the help of sense. This is presented to us in a dreamy manner, and yet is said to be necessary, for we say that all things must be somewhere in space.…To sum up: Being and generation and space, these three existed before the heavens, and the nurse or vessel of generation, moistened by water and inflamed by fire, and taking the forms of air and earth, assumed various shapes.[44]

The space Newton inherited, the receptacle space in the universe God preformed, did not need its origins explained nor did it originate anything. Because receptacle space and the objects in it did not include any notion of generation and change, this is the one Whitehead urged scientists leave behind.[45] For his part, Newton deferred the problem of determining the nature and causes of forces in favor of developing laws for predicting motion.[46] Further, Whitehead repudiated "the doctrine of vacuous actuality" and "the distrust of speculative philosophy" while scientists read Newton as urging them to forego hypothesizing.[47] Scientists focused on observing, identifying, and quantifying the forces of nature as seen in moving objects. They used empirical observations to develop formulas and tested them by making

[40] *Ibid*, p. xi.

[41] *Ibid*, p. 209.

[42] *Ibid*, p. 123.

[43] Lukerman 1961, p. 196.

[44] Plato's *Timaeus*, 1871/1953, pp. 649–650.

[45] "Receptacle" The Concise Oxford Dictionary of English Etymology. Ed. T. F. Hoad, 1996.

[46] Despite his methodology, Newton never altogether discarded the goal of understanding the nature and causes of forces such as gravity. Newton, 1999, p. 588 (italics added), "I use the word "attraction" here in a general sense for any endeavor whatever of bodies to approach one another… Mathematics requires an investigation of those quantities of forces and their proportions that follow from any conditions that may be supposed…these proportions must be compared with the phenomena, so that it may be found out which conditions of forces apply to each kind of attracting bodies. *And then, finally, it will be possible to argue more securely concerning the physical species, physical causes, and physical proportions of these forces*."

[47] Whitehead 1929a, b, p. xiii.

predictions. Most famously, many assume Newton urged scientists to avoid hypothesizing about possible causes of forces such as gravity. Instead, Newton argue they should be satisfied with what one can be certain of through observation and deduction.

> I have not as yet been able to deduce from phenomena the reason for these properties of gravity, and I do not feign hypotheses. For whatever is not deduced from the phenomena must be called a hypothesis; and hypotheses, whether metaphysical or physical, or based on occult qualities, or mechanical, have no place in experimental philosophy. … it is enough that gravity should really exist and should act according to the laws that we have set forth and should suffice for all the motions of the heavenly bodies and of our sea.[48]

Given this recurrence to Plato and Newton, among other philosophers, it is difficult to overestimate the relationship between understanding the history of science and understanding Whitehead's natural philosophy and view of conscious experience. For example, when Whitehead taught his philosophy of science classes at Harvard, he reviewed Newton and the history of traditional thinking about space, time, and material.[49] Newton proposed absolute space as a stationary stable 'background' against which the motion of objects would be apparent. Yet, when objects are in absolute space, in absolute motion relative to each other, and each exhibits constant velocity, it is impossible to tell if any one object is resting or moving. When scientists began to think of space as relative space, simply the relationship among bits of material, most still implicitly also conceived of it as separate from material. This implicit assumption manifested as a tendency to speak of objects *in* space when there was purportedly no space for them to be 'in.'

Unfortunately, relative space made it difficult to explain phenomena such as earth's equatorial bulge or the direction of cyclonic rotation. In order to make sensible measurements (e.g., of movement, distance, and time in space) something must be fixed and unchanging as the point of comparison. If the earth's bulge and cyclonic rotation are not made apparent relative to absolute space, and relative space does not exist independently, what could be the stable framework revealing these effects? Not just Newton, but Whitehead and many scientists have suggested some form of aether, though not perceptible with modern instruments, may be the material making motion apparent. Tests, however, failed to support the aether hypothesis.[50]

Whitehead noted Einstein's brilliance in the realization that while a single entity may not exhibit non-uniform motion in relation to other single entities, a *group* of entities could exhibit non-uniform motion relative to another *group*. This helped solve the absolute motion problem of two observers on earth by treating each observer as a group of molecules rather than individual objects.[51] Einstein could relate each observer's measurement of space to that observer's measurement of time using the velocity of light (c) as a constant for determining a simultaneous starting point.

---

48 Newton 1999, p. 943.

49 Marvin 1927–1928, p. 10.

50 Whitehead 1922, p. 5.

51 Marvin 1927–1928, p. 160.

Whitehead was still concerned, however, about the tendency among scientists to assume implicitly that relative space could be a receptacle for objects in space when theoretically no such space existed.[52] He also suggested there is a problem with calling *c* a 'constant' because to derive light's velocity as an invariant Einstein modified Maxwell's equation to account for the effects of gravity and used the speed of light in a vacuum. Since in nature light does not occupy a vacuum, else why would gravity alter its velocity, Whitehead saw grounds for disputing this interpretation.[53] Finally, Whitehead expressed a difficulty he saw with general relativity. Both a stable framework and a foundation for congruence of some kind are necessary for measurement to make sense. Whitehead explained, "But Einstein's interpretation of his procedure postulates measurement in heterogeneous physical space, and I am very sceptical [*sic*] as to whether any real meaning can be attached to such a concept."[54]

For Whitehead all three scientific concepts, space, time, and inert material, were abstractions from events labeled scientific objects.[55] Whitehead appreciated Minkowski's idea of particle occurrences[56] and argued that an instantaneous point is, "better named an 'event particle'" because the concept of precise insubstantial points in precise time was an abstraction.[57] With respect to gravity, Whitehead agreed with Einstein's basic formula and methods for transforming measurement from one space-time to another. He did not, however, see the grounds for assuming disconnected particles that only change the properties of space time for nearby particles. Such an arrangement implied the existence of empty space. When he compared his theory of gravity to Einstein's he stated, "But the essence of the divergence of the two methods lies in the fact that my law of gravitation is not expressed as the vanishing of an invariant expression …".[58]

Instead, Whitehead used the same tensors Einstein used[59] but calculated 'impetus' by using two functions, the 'gravitational potential', and the 'associate potential.' He calculated the associate potential using direct rather than inverse distance and this meant it did not vanish between particles.[60] In *Principles of Relativity* he described his view of aether as active and not a 'shy agent behind a veil.' To be more precise, Whitehead (1922, p. 37) did not believe it was simply something in which other things existed but stated, "In the classical doctrine the ether is the shy agent behind the veil: in the account given here the ether is exactly the apparent world, neither more nor less." Whitehead described physical nature,

[52] Whitehead 1916/1961, p. 93.

[53] Whitehead 1948/1961, p. 127.

[54] Whitehead 1948, p. 312.

[55] Carr et al. 1922, p. 128.

[56] Minkowski 1909/2009.

[57] Whitehead et al. 1919, p. 33.

[58] Whitehead 1948, p. 313.

[59] A tensor is a mathematical entity with components that change in a particular way in a transformation from one coordinate system to another.

[60] *Ibid*, p. 306.

> The physical properties of nature arise from the fact that events are not merely colourless things which happen and are gone. Each event has a character of its own. This character is analysable [*sic*] in two components:—(1) here are the objects situated in the event; and (2) there is a field of activity of the event which regulates the transference of the objects situated in it to situations in subsequent events…. Space and time have their origin in the relations between events. What we observe in nature are the situations of objects in events. … The whole complex of events viewed in connexion [*sic*] with their characters of activity takes the place of the material ether of the science of the last century. We may call it the ether of events.[61]

Thus, the idea of an extensive continuum was built into Whitehead's gravitational formula and I argue this renders his space much closer to the generative space, the *choros*, of Plato's *Timaeus* and perhaps to modern quantum theoretical views. Whitehead's former student, A. S. Eddington reported that Whitehead's equations gave the same answers as Einstein's except when considering, " … more than one attracting particle…" because Whitehead allowed "… superpositions."[62] The Einstein/Whitehead gravity comparison is still controversial, but modern authors may not take into account the relevance of Eddington's 1924 observation.[63]

Figure 13.1, reproduced from Whitehead's *The Principle of Relativity* depicts Whitehead's idea of an event-particle's historical route.[64] As he explained, the diagram of an event particle exhibits the fundamental properties of the extensive continuum of the universe. In general, "This structure is four-dimensional, so that any event is a four-dimensional hyper-volume in which time is the fourth dimension. But we should not conceive an event as space and time, but as a unit from which space and time are abstracts."[65] This analogy is limited, but one way to picture what Whitehead meant by an extensive continuum of events was that unlike discrete particles with limited fields event-particles exhibit more 'spread.' An event particle's 'spread' is its historical route in space-time. Rather than thinking of nature as comprised of discrete particles forming more or less enduring unions into atoms and molecules that move around together in empty space, there is no empty space. Imagine nature as comprised fundamentally of change. Event particles constantly incorporate data provided by the event particles in their past. Event particles exist as evolving processes not as discrete entities physically separate from each other or separate from what we perceive as empty space. They have both a rhythmic and a vector like quality because similarity contributes to similarity but newness, no matter

[61] Whitehead 1948, p. 311. This is very similar to the influential point atom that Roger Boscovich described in 1763. Boscovich described a theory of space and material in which atoms were merely points of force. For instance, Johann Gottfried Herder used this concept in his philosophy as *Krafte* (force). See Edwards (2013) and Zammito (1992) for some implications of this physical theory for eighteenth century German philosophy.

[62] Eddington 1924, p. 192.

[63] See, for example, "On the multiple deaths of Whitehead's theory of gravity" by Gibbons and Will (2008) and "Whitehead contra Einstein" by Reinhardt and Rosenblum (1974). That physicists still examine Whitehead's theory of gravity lends it more credibility than many may have assumed it deserved.

[64] Whitehead 1922, p. 31.

[65] *Ibid*, p. 33.

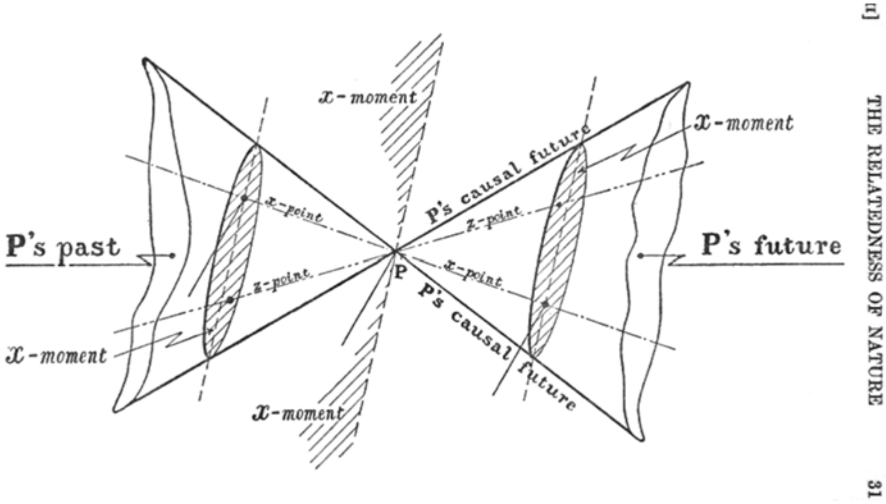


Fig. 13.1 Depiction of the historical route of a particle. Whitehead noted that he borrowed the term 'historical' from Professor C.D. Broad (Reproduced from Whitehead 1922, p. 31)

how slight, is inherent. The spread of an event particle includes the wake of its past, the present 'final actual occasion,' and its forward spread into the next event it will become for a perceiver Because change is a fundamental characteristic of nature, the events are the relata of the extensive continuum and not disconnected enduring objects doing things in empty space over time. At the most microphysical level events include producing, including, and excluding potentials, and producing valuated probabilities as depicted in Schrödinger's equation for quantum mechanical state evolution. As part of the overall event, potentials have ontological significance but are not to be conflated with actualities.

As Whitehead described the concepts of causal past and causal future, what must have seemed clear to him from the observations quantum physicists reported was that nature is inherently continuous change but that it must have both a yes/no decision quality combined with its quality of continuity into a new phase of change. If nature were not always productive of a yes/no decision quality, there would be no objectively real 'objects' or 'relative space' and if nature were not always producing something at least a little new, there would be no production of time. The causal past referred to nature's physical pole, the final real actualities carrying efficient causation and determinism. They were the objective data entering into what could become next. Because all events do not occur.

It may clarify what this means in terms of consciousness to note that in 1922 Whitehead explained that when he referred to a perceptual object as a pervasive adjective of an event of the extensive continuum he meant that it was a "true Aristotelian adjective" meaning "an adjective of any temporal slice of that event." They condition experienced events,

> … a perceptual object means a present focus and a field of force streaming out into the future. This field of force represents the type of control of the future exercised by the

> perceptual object—which is, in fact, the perceptual object in its relation to the future, while the present focus is the perceptual object in its relation to the present. But the present has also a duration. What we observe is the control in action during the specious present.[66]

Whitehead elaborated some of his ideas for a philosophy of science class in 1927. The causal future, the nextness, is the potentia that form from the merging of objective data with subjective data. The potentia will become the valuated probabilities, only one of which will finally occur as the final real yes/no. The causal part of the term 'causal future' the fact that the potentia and the valuated probabilities are conditioned by the objective and subjective data; not every outcome is possible and among those that are possible, not all are equally probable.[67]

The future part of the 'causal future' referred to the newness, no matter how slight, that always appears. By causal future he was not referring to any kind of mystical reaching backward in time from a decided future to an undecided past. Instead he referred to the idea gleaned from quantum physics that there is a period inherent in the continuity of process when potentia must be counted as physically real without being decided because these potentia impose real limits on and contribute to the future. Without them nothing would exist. Nature included, for Whitehead, an actual finally real past of determined objective events, a determined and conditioned specious present of causally independent contemporary events, and indeterminate but conditionable potentia, only one of which will become finally real.[68]

When Whitehead referred to his philosophy as a philosophy of organism, he was not in any sense trying to superimpose the idea of animate entities onto non-living entities in an arbitrary manner.[69] He argued that,

> The status of life in nature is the standing problem of philosophy and of science. Indeed it is the central meeting point of all the strains of systematic thought, humanistic, naturalistic, and philosophic. … neither physical nature nor life can be understood unless we fuse them together as essential factors in the composition of "really real" things whose interconnections and individual characters constitute the universe.[70]

He attributed the problem of explaining life from inert material to the Cartesian distinction between material and mental substances as independent of each other. Eventually in the thinking of European scientists, living and mental were fused and non-living and physical material were fused. "The effect of this sharp division between nature and life has poisoned all subsequent philosophy."[71] He developed his interpretation of physical events as inherently possessing characteristics typically ascribed to animate entities for two reasons. First, thinking of change as the stopping and starting of movement in a vacuous receptacle space occasionally occupied by disconnected inert material made change logically impossible. He saw

[66] *Op cit*.

[67] Marvin 1927–1928, pp. 56–57; See also Whitehead's diagram and explanation on p. 80 of this text.

[68] Whitehead 1929a, b, pp. 319–320.

[69] *Ibid*, p. 18.

[70] Whitehead 1934, p. 3.

[71] *Ibid*, p. 4.

no way around taking the change/movement so characteristic of life as a given. Further, quantum physics gave every reason to propose that at the microphysical level the universe was inherently changing and moving. Therefore, he speculated that the way out of incoherence lay in reversing the scientific view of nature from one that explained experiences of a dynamic system by reference to inherently inert parts in a vacuous reality into one that explained experiences of inert parts by reference to an inherently dynamic system of events.

This brief synopsis does not do justice to Whitehead's analysis. The main idea was to show that Whitehead derived his event ontology by critically examining ideas in the history of science and their influence on observations and theorizing in contemporary physics. From there he formulated and tested a natural philosophy consistent with his interpretation. Weber recently likened Whitehead to a ghost whose work, "… still remains so to speak in scholarship limbo…" because he has had relatively little impact on mainstream science.[72] A renewed interest in Whitehead is emerging,[73] however, and Epperson argued that Whitehead's view of nature is congruent with current quantum theories such as those of Gell-Mann and Hartle 2007, Griffiths and Omnès 1999, and Zurek 2002.[74] According to Whitehead, the difficulties inherent in scientific materialism emerged because, unaware, scientists inherited assumptions from earlier thinkers that led them to superimpose sharp physical and categorical separations on their definition of "material" and what is really an extensive continuum.

13.3 Chalmers, Whitehead, and the Hard Problem of Conscious Experience

When Chalmers introduced the distinction between the easy and the hard problem of consciousness he focused exclusively on accounting for conscious experience.[75] This was unlike Whitehead, whose focus was not on consciousness but on developing a speculative philosophy and general scheme in which consciousness and (everything else of which we are conscious) would be a particular instance. Similar to Whitehead, Chalmers turned to physics but Chalmers used physics to put his argument in context when he compared the study of experience to the study of physics. Later, Chalmers' takes quantum theorists' ideas into account but by his own admission has not changed his views substantially and still upholds essentially the same position he outlined when he first described the hard problem.[76]

[72] Weber 2006, p. 118.

[73] For example, Weber and Weekes 2009.

[74] Epperson 2004, p. 53.

[75] Chalmers 1995.

[76] Chalmers 2010, p. 15.

Chalmers' work as whole is properly credited with focusing attention on the conscious experience of qualia by providing a particularly lucid and accessible account of the key features that have made it such an intriguing and intractable problem for so long. He demarcated the hard problem of conscious experience from the easier problem of simple awareness because the easy problems of consciousness are functions that accompany observable behavior.[77] Experience as he defined it is equated with terms such as qualia and with the 'something it is like' to be in a state. Chalmers has faith that a reductive approach will sufficiently explain the existence of living systems and their functions such as cognition, memory, access to internal states, verbal reports, language, and perception. Chalmers explains, however, that, "By contrast, the hard problem is hard precisely because it is not a problem about the performance of functions. The problem persists even when the performance of all the relevant functions is explained."[78]

Individuals who exhibit cortical blindness (or blindsight) following brain injury provide one potentially useful example of what Chalmers and others may mean by conscious experience as opposed to functional awareness. Cortical blindness may manifest as a modality specific difficulty recognizing objects and may depend on an intact lateral geniculate nucleus despite damage to striate cortex, for example.[79] Individuals with blindsight give verbal reports indicating they do not have visual experiences and from that standpoint are sightless although they can recognize objects by other sensory modalities such as touch. Thus, depending on the nature and severity of damage, individuals with blindsight report relatively little if any conscious experience of qualia (the hard problem of consciousness) but demonstrate awareness (an easy problem of consciousness) because they are able to respond fairly accurately when required to point or otherwise indicate the locations of objects in what would be their visual field.[80]

Chalmers proposed that understanding conscious experiences of qualia requires an extra ingredient, a non-reductive explanation, not to be found in any of the physico-reductive schemes.[81] Ultimately, he asserted that conscious experience should be taken as a given in nature, an ontological addition in the same spirit that physicists take fundamental entities such as mass and space-time as givens and proceed to build theory and explanation from there. For a satisfactory theory of consciousness, Chalmers argued that one needs to find which processes give rise to experience and an account of why and how. He also suggested taking as given that anything in conscious experience will also be represented cognitively. This connected awareness with experience and made it feasible to argue that neural substrates for simple awareness will also partly explain conscious experience. In addition he suggested taking as given that the patterns of interaction among causal

[77] Chalmers 2010, pp. 5–10.

[78] Chalmers 1995, p. 203.

[79] Sahraiea et al. 2011, pp. 21217–21222.

[80] Schmid et al. 2010, p. 374.

[81] Chalmers 1995, p. 204.

components and not the specific physical make-up of those components contribute to the experiential component of consciousness.[82]

With the aforementioned constraining principles in place, Chalmers then proposed that an information theory in which all information has a physical aspect and a phenomenal aspect might underpin these two constraining principles and conscious experience itself. Information is physically embodied and "… has a basic structure of difference relations between its elements."[83] In arguing for a connection between information and conscious experience, he noted that physics deals primarily with the extrinsic properties of entities and that it is possible that if intrinsic physical properties exist, "We might say that phenomenal properties are the internal aspect of information."[84]

Chalmers admits to the speculative nature of his ideas and to the fact that it would be necessary to determine if all information has a phenomenal aspect or if experience is only to be found with certain sorts of information.[85] Finally he suggests that there need be no such constraint and that experience could easily be a widespread physical phenomenon. Simple information processing, such as that in a mouse or a thermostat, gives rise to simple experience and complex information processing gives rise to complex experiences. Both animate and inanimate entities can be thought of as experiencing,

> … experience is much more widespread than we might have believed … on reflection I think the position gains a certain plausibility and elegance … Indeed, if experience is truly a fundamental property, it would be surprising for it to arise only every now and then; most fundamental properties are more evenly spread.[86]

From his perspective, the physical domain is causally closed but conscious experience, as the product of information processing, may supplement the physical domain. Therefore, understanding experience may allow scientists a way to impart a causal role to our experiences of qualia.

Important differences between Whitehead's and Chalmers' views almost immediately appear. Whitehead rejected materialism and dualism. Chalmers accepted purely reductive physical explanations for cognitive functions but proposed conscious experience requires a non-reductive explanation in the form of a dualist information theory such that all information has a physical part and a phenomenal part. It is difficult to tell if Chalmers meant that all that exists physically is information and its physical and phenomenal parts are mutually implicative but the fact that he finds purely reductive explanations satisfactory for the easy problems and only turns to the non-reductive for the hard problem suggests this may not be so. From Whitehead's perspective, the problems exemplified by the hard problem of consciousness are due to incoherence, indicated by one kind of explanation for some

[82] *Ibid*, pp. 218–220.
[83] *Ibid*, p. 219.
[84] *Ibid*, p. 218.
[85] *Ibid*, p. 217.
[86] *Ibid*, p. 215.

phenomena with other types of explanation for other phenomena. In addition, incoherence is seen in bifurcating nature into mutually exclusive categories such as inanimate/animate and material/immaterial, all of which Whitehead recast as continua or mutually implicative. Despite what appear to be deeply divided views of nature, however, comparing Whitehead's and Chalmers' views of the hard problem reveals several marked similarities in the two men's approaches.

The issue of vocabulary is particularly germane here because Whitehead sought a coherent natural philosophy without explanatory gaps. Whitehead stated that "The modern natural philosophy is shot through and through with the fallacy of bifurcation … Accordingly all its technical terms in some subtle way presuppose a misunderstanding of my thesis."[87] In seeking a set of general ideas applicable to everything in nature Whitehead concluded that traditional scientific language was problematic because it developed under the assumption that certain categorical distinctions reflected real differences in nature. Consequently, scientists and philosophers used different vocabularies for investigating human, other animate, and inanimate systems. Therefore, comparing Whitehead and Chalmers on the hard problem of conscious experience will entail keeping the thread of Whitehead's natural philosophy firmly in hand throughout so that his meaning is not misconstrued by applying modern definitions.

Keeping in mind that the two men were not focused on the same goals, it is necessary to choose some bases for comparing their views of consciousness. These points for comparison are by no means exhaustive and will lead to some overlap but they highlight some of the common features between the earlier and later thinking about consciousness. The subsequent focus therefore will be on: where each stood with respect to framing scientific questions about consciousness, the relationship between consciousness and experience, the relationship between subjects and objects, and their views of cause.

13.4 Framing the Question

In the final analysis, Chalmers argued that scientific analysis of conscious experience will require an extra ingredient because experiences are fundamentally distinct from material. It is for this reason that he adds experience to the traditional ontology of physics. Like material and space-time in the physical universe, it is a fundamental given in nature. Therefore, he frames the question about the science of consciousness not as explaining the existence of consciousness but as requiring a theory that accounts for its properties. He does not argue that conscious experiences are necessarily epiphenomenal but suggests they may be epiphenomena or they may have a subtle causal role.[88] In this view, conscious experiences are not explicitly rejected

[87] Whitehead 1920/1955, p. vi.

[88] It is difficult to determine what Chalmers sees as the difference between a causal role and a subtle causal role.

from nature but the assumption is that they are immaterial, internal, subjective phenomena that have an individuality not found in third-person data.

Obviously, Whitehead might first argue that Chalmers' assumption that a different explanation is sufficient for the easy and the hard problems reflects an inconsistent account of nature. Whitehead discussed the necessity of viewing all natural events as gradients, including the gradients between internal/external, organism/environment, and animate and inanimate. He took a strong stance against anything in nature as epiphenomenal stating,

> The effect of this sharp division between nature and life has poisoned all subsequent philosophy. Even when the co-ordinate existence of the two types of actualities is abandoned, there is no proper fusion of the two … For some, nature is mere appearance … For others physical nature is the sole reality and mind is an epiphenomenon. Here the phrases "mere appearance" and "epiphenomenon" … carry the implication of slight importance for the understanding of the final nature of things. The doctrine I am maintaining is that neither physical nature nor life can be understood unless we fuse them together as essential factors in the composition of really real things …[89]

Whitehead would not accuse Chalmers of dismissing mind from nature but he might argue that Chalmers bifurcates nature because he opts for naturalistic dualism. Chalmers adds experience onto material as a new non-physical ontological category of nature. Thus, rather than experience as intrinsic to nature, it can only be understood by developing bridging principles between materials and experience. In this view conscious experiences are still immaterial as opposed to physical structures and processes. Compared to Whitehead's process ontology, the material or physical side of Chalmers' nature bears a disproportionate amount of causal weight although he allows for the possibility that experience may have subtle effects.

Whitehead, like Chalmers, keeps experience as fundamental and views it as existing along a gradient of complexity from simple inanimate to complex animate. Unlike Chalmers, however, Whitehead's event ontology does not require bridging principles between physical features of nature and non-physical experience. In Whitehead's natural philosophy all fundamentals apply to every type of event so that there is no division between animate material and inanimate material, or between material and immaterial, it is all process. Whitehead accomplished this by defining experience as any effect of one event on another so it can be as simple as the absorption of energy. Both Chalmers and Whitehead emphasize an information component to experience in nature. Whitehead used the term data but Whitehead's nature is comprised of mutually implicative events that constantly evolve by incorporating data and differ primarily in complexity. Causes inform effects in any event by entering into the events in process.

Further, Chalmers postulated one kind of reductive explanation for the easy problems and a different kind of bridging or relational explanation for the hard problem of conscious experience. Whitehead's event ontology assumed that any question in science, including a question about the experience of redness is a question of how one natural event is related to other natural events,

[89] Whitehead 1934, pp. 4–5.

> … the fire is burning and we a see a red coal. This is explained in science by radiant energy from the coal entering our eyes. But in seeking for such an explanation we are not asking what are the occurrences which are fitted to cause a mind to see red. The chain of causation is entirely different. The mind is cut out altogether. The real question is, when red is found in nature, what else is found there also? Namely, we are asking for an analysis of the accompaniments in nature of the discovery of red in nature.[90]

Whitehead agreed with Chalmers' decision to keep conscious experience in nature as real and worthy of explanation. He would likely agree with Chalmers' framing the study of consciousness as a quest for principles of relatedness. Only Whitehead might add that this ought to be the same for all questions about nature not just the hard problem of conscious experience.

13.5 The Relationship Between Consciousness and Experience

Any comparison between Whitehead and Chalmers would do well to examine their definitions of consciousness and experience. Chalmers differentiated between two kinds of consciousness. Consciousness as only awareness and conscious experience. For Chalmers experience is the qualitative aspect of consciousness so that experiencing qualia presupposes awareness. One could, however, behave under the control of information (be aware) without necessarily having conscious experiences of qualia. This is the situation that one finds in the cortical blindness example described previously. These individuals behave in a way that demonstrates simple awareness of the location of objects, for example, because they can respond differentially. But they report that they are blind, that they are not experiencing any visual qualia.

Whitehead used the term experience differently. Like Chalmers, experience for Whitehead referred to something fundamental in which all entities engaged. Although Chalmers calls it information processing, Whitehead referred to it as a necessary process of all events that involved incorporating data from the environment stating, "The principle that I am adopting is that consciousness presupposes experience, and not experience consciousness."[91] Chalmers' view implies that the information/experience and the experiencing subject have a kind of functional independence from each other. The experience is in or by the subject without suggesting any further change beyond location of the information. Whitehead proposed that experience was a process whereby any event (entity, electron, other event particle, event-object) incorporates data provided by other events into its process of becoming what it will be next.[92] Chalmers' view implies the subject has experiences or takes in information so that it is fundamentally the same enduring subject with the addition of this experience which might then alter what that same subject will do.

[90] Whitehead 1920/1955, p. 39.

[91] Whitehead 1929a, b, p. 53.

[92] Marvin 1927–1928, p. 17.

For Whitehead, the experiencing subject is incorporating on the way to becoming something else.

In addition, while both Chalmers and Whitehead offer a broad view of what constitutes an experiencing subject, since Whitehead defined all of the terms in his natural philosophy as generally as possible experience is the fundamental generative mode of all of nature. Without it, nature would cease to exist because there would be no process. All actual entities that most people typically classify as animate or inanimate are events or societies of events in Whitehead's ontology and are necessarily centers of 'experience.' Any event that evokes activity in another is an 'object' and the event in which activity is evoked is a prehending, experiencing subject. Experience, Whitehead told his class, "… is what Physics deals with in a highly abstract way. Psychology also."[93]

Further, Whitehead emphasized a kind of reciprocity that is absent from Chalmers view. Any event (amoeba, rock, or a person, for example) that is experiencing is necessarily providing some of the data and therefore is always part of the new data evoked by an object-event. Whitehead also used the term 'feeling' in a way that is divested of animate/inanimate distinctions. The data are felt, but objectively, in the sense that they evoke activity. The objective data necessarily must also become 'subjective,' however, because objective data becomes part of the experiencing event. Data provided by the object-event may be selected by inclusion or selected for exclusion. For example, ignoring something can have an important effect on an outcome just as much as including it. Whitehead stated that "In awareness actuality, as a process in fact, is integrated with the potentialities which illustrate *either* what it is and might not be, *or* what it is not and might be. In other words, there is no consciousness without reference to definiteness, affirmation, and negation."[94]

As will be discussed further in the next section, Whitehead went further in his natural philosophy and explicitly separated the subject/object structure of experience from the knower/known relationship.[95] All experience is structured as subject/object relationships but knower and known are highly abstract and only apply in complex entities. Since Whitehead divested the subject/object structure of experience from the knower/known, the roles of subject and object become entirely relative. One does not have to ask about differences in complexity. The subject is not required to be capable of knowing or being aware and the object is not required to be inside or outside the subject or to have any particular relationship to a mind. Objects provide data that affect subjects but the effect is a subjective interrelating process, not necessarily an animate entity knowing something or awareness of an event external to it.

---

[93] Marvin 1927–1928, p. 68.

[94] Whitehead 1929a, b, p. 243. Note that I suspect that the first comma in the first sentence of this quote was misplaced. I think the intent was to say that 'actuality is integrated with potentialities' so it should read "In awareness, actuality … is integrated with the potentialities …" As the editors of the corrected edition of *Process and Reality* noted, Whitehead was not known to enjoy correcting proofs.

[95] Whitehead 1931/1961, pp. 222–239.

Whitehead also used the term prehension to make concepts such as experience and consciousness more general to all of nature and not a special concept for people. Prehensions are processes of appropriating data, a temporary category of existence that occurs during events which Whitehead also referred to as 'occasions of experience.'[96] Prehensions do not require a knower/known relationship so rocks can prehend in the sense that they can become something different by absorbing energy. Prehensions can be positive or negative. For instance, Schrödinger's equation for calculating the outcome of quantum mechanical state evolution involves the production of a matrix of potentia that includes possible outcomes and even contradictory outcome states. These potentia, however, evolve into a matrix of valuated probabilities by a process of cancellation so that contradictions are eliminated. Only one of the probable outcomes actually occurs. Similarly, positive prehensions are events that will be further included in the prehending subject-event as a relevant, possible emphasis for the subject. Negative prehensions might be thought of as potentia that are cancelled out and relegated to the background of vague existence.[97]

Like Chalmers, Whitehead also believed that for us to understand 'experience' we may need to call upon an additional ingredient although the two men used the term 'experience' differently. In discussing the explanation of conscious experience Chalmers asserted that the extra ingredient needed is a non-reductive explanation for experiences of qualia in the form of an information theory and bridging principles. Consciousness is awareness and conscious experience is special, it is the experience of qualia for Chalmers that is the hard problem. Whitehead argued that for understanding the fundamental experience process in nature consciousness itself is a possible extra 'ingredient' of experience.[98] When we are experiencing (which is all the time) we may also experience at a higher grade that includes consciousness. "Experience is wider than consciousness. We are aware of focus of attention, but also know that a mass of experience is in the background, unanalyzed."[99] Also similar to Chalmers, for Whitehead, inanimate objects experience, i.e. have experiences. Whitehead makes a complexity argument similar to Chalmers' in that not all entities are complex enough to experience 'consciousness' the way people use this term about other people.

Specifically, for Whitehead experience and consciousness occur as gradients from, for example, experience in terms of microphysical activity in an event particle provoked by another event particle to the provocation of complex conscious experiences of qualia in a person eating a pizza. In addition, beyond a grade of experience, consciousness was for Whitehead a temporary shift in emphasis on or 'concern for' one aspect of all nature while simultaneously de-emphasizing the rest. The aspect of nature experienced could be oneself, but the rest of nature must always be in the background as the contrast against which the aspect prehended stands out.[100]

[96] Whitehead 1934, p. 16.

[97] Marvin 1927–1928, p. 70.

[98] For example, Marvin 1927–1928, p. 17.

[99] *Ibid*, p. 36.

[100] Whitehead 1931/1961, p. 223.

For example, because existence in Whitehead's ontology necessarily entails the constant absorption of data and these data necessarily include the data provided by the entity itself, part of experiencing something as enduring, such as an unchanging blue sky, is the person's immediate past experiences, no matter how vague, of blueness and 'otherness not blue.' "Whenever there is consciousness there is some element of recollection. It recalls earlier phases from the dim recesses of the unconscious. Long ago this truth was asserted in Plato's doctrine of reminiscence."[101] Events appear to endure unchanged because similar antecedent events consistently evolve into similar 'specious present' events. For Chalmers, mere awareness of brown could be an example of the discriminative, functional, objective awareness one could easily account for with a reductive explanation and the conscious experience of brown qualia was the difficult to explain subjective event.

As Chalmers distinguishes between functional awareness and experience of qualia, Whitehead also distinguished between grades of experience but suggested that the extra ingredient consciousness adds to experience is the contrast between awareness of what could be but is not while simultaneously maintaining awareness of what is. Whitehead described this for his philosophy class,

> When I feel a thing as not grey I am at the height of consciousness. When I feel it merely as brown (which it is) the concept hasn't a heightened intensity of feeling that comes from contrast. There is always some element of the negative in concepts and consciousness.[102]

For Chalmers everything that is consciously experienced is also cognitively experienced.[103] Whitehead's view suggests that the experience of qualia, the private sense of "what it feels like to be experiencing redness" is an experience that results from the objective data being incorporated into the subjective data. It includes both the inclusion process and the exclusion process inherent in natural events. Furthermore, because Whitehead views change and activity as already inherent in nature, he does not need to make a sharp distinction between inside and outside, events that start behavior and stop behavior, or real events and representations of events. This also breaks down sharp distinctions between subjective and objective, experience and consciousness, and conscious and unconscious. He only has to refer to gradients of complexity and differences in the direction or vector of change already ongoing.

Finally, both Whitehead and Chalmers refer to externality and internality in discussing the ideas of consciousness and experience although they approach these from different angles. Whitehead did not believe any sharp internal/external distinction occurred and did not see such sharp distinctions as particularly helpful to understanding nature or consciousness. He viewed the internal subjective aspect of experience (broadly defined as incorporating data) as the 'mental pole' and intrinsic to the process of becoming when objectively provided data are merged with the subjectively provided data for the generation of potentia. Intrinsic to the process did

---

[101] Whitehead 1929a, b, p. 242.

[102] Marvin 1927–1928, p. 69.

[103] Chalmers 1995, p. 19.

not necessitate internal to the body of an animal, though. Via contrasts inherent in this process, adversions and aversions may occur. In complex entities such as people these may make some outcomes more or less likely. Thus, the contrasts could play an important role for the generation of qualia that may then have a causal role in choices.[104] In Whitehead's philosophy, part of the present must and does contain some consideration of the future. In his *Louis Clark Vanuxem Foundation Lectures* at Princeton he made plain the difficulty inherent in a science that refuses to acknowledge the role of foresight in understanding some operations of animal bodies.[105]

Chalmers suggested something similar about the intrinsic physical qualities as informational, "Once a fundamental link between information and experience is on the table, the door is opened to some grander metaphysical speculation concerning the nature of the world. …it is often noted that physics characterizes its basic entities only extrinsically, …If one allows that intrinsic properties exist …We might say that phenomenal properties are the internal aspect of information."[106]

13.6 The Subjective/Objective Distinction

Chalmers represented the difference between problems that are and are not amenable to reductive models as requiring objective and subjective data, respectively. Implicit in this methodological problem is a distinction between the 'aboutness' of these kinds of data which is loosely tied to the location of the information. Objective data are about events outside the privacy of conscious experiences while subjective data are problematically private and about something inside. Whitehead reframed subjective/objective perspectives on experience as mutually implicative events because he saw difficulties arise when one assumed that human-centered thinking implicitly assumed all experience is about nature rather than assuming it is included in nature. He intentionally divested the subject-object dichotomy of its human-centeredness.[107] In addition, most assume that simple and fundamental elements of experience necessarily are the most clear, exact examples of experience. For example, one assumed an experience of 'redness' would be a very simple experience. This assumption, Whitehead contended, is false because it confounds the term 'subject' with 'knower' or 'knowledge' and the term 'object' with 'known.'[108]

[104] Whitehead 1929a, p. 234 and 241, viewed adversion and aversion as a valuation up or down, respectively, and the valuation was a combination of importance and intensity.

[105] Whitehead 1929b, *The Function of Reason.*

[106] Chalmers 1995, p. 219. It is difficult to tell but the supposition that physicists in 1995 characterized basic entities only extrinsically and that 'if one allows that intrinsic properties exist' may reflect a misunderstanding of physics and its history. Since the early quantum experiments of the 1920s and later few physicists saw the physical world this way. See for example Merzbacher (2002) on the long history of quantum tunneling.

[107] Whitehead 1931/1961, pp. 223–239.

[108] *Ibid*, pp. 223–224. See also Dewey and Bentley (1949). *Knowing and the known*.

Accordingly, as mentioned above, for Whitehead subject and object in the subject-object relationship were not only mutually implicative but entirely relative. For Whitehead, an 'object' was any event that provoked activity in another event. A subject was any event having activity concerning an event-object. In this situation, most of the time for every experiencing subject there could easily be a reciprocal relationship with an experience-evoking object. The distinctiveness, vagueness, or qualitative nature of experience was not necessarily an indicator of subjectivity or objectivity. For Chalmers the bifurcated nature of the data, information about something inside a subject or about something outside a subject, drives the bifurcation of explanation while for Whitehead there is a continuum of subjectivity to objectivity and one kind of explanatory approach, relational.

The assumption of distinctions between objectivity and subjectivity, particularly in cases where other dichotomies such as mind/brain and organism/environment are assumed, maintains the barrier that Chalmers and others work to overcome in developing theories of conscious experiences of qualia. As Tiebout pointed out, however, Whitehead did not think that questions about the difference between nature as a whole and nature the way it appears to an individual is bifurcating nature. He did argue that none of the data available for understanding nature is exclusively 'in the mind' or 'out of the mind.'[109] Therefore, Whitehead's view of the relativity of subjective and objective is consistent with Chalmers' argument for the validity of using verbal reports from research participants as first-person data.[110]

Both Chalmers and Whitehead argue that similar organization is likely to provide similar experiences. Chalmers makes this point via the principle of organizational invariance, "… what matters for the emergence of experience is not the specific physical makeup of a system, but the abstract pattern of causal interaction between its components."[111] Whitehead took this stance by allowing a complexity gradient to account for differences in experiencing events. Whitehead explicitly extended ideas about subjective, objective, and qualia into the question of a relative relationship between subject and object. Similarly, Chalmers approaches the relationship between scientists and nature stating,

> … analysis of the cognitive explanation of our judgments and claims about conscious experience—judgments that are functionally explainable but nevertheless deeply tied to experience itself—suggests that explanation centrally involves the information states embedded in cognitive processing. It follows that a theory based on information allows a deep coherence between the explanation of experience and the explanation of our judgments and claims about it.[112]

One implication is that realistically, whether one is considering objective data as a scientific method or objective data as a person decides what to have for dinner, it is difficult to uphold an absolute distinction between objectivity and subjectivity. In science this is because people making observations always bring a different

109 Tiebout 1958, p. 45.

110 Chalmers 2010, p. 48.

111 Chalmers 1995, pp. 218–219.

112 *Ibid*, p. 218.

physical perspective if not different experiences and interpretations. It is less difficult, and possibly closer to the human component of the scientific enterprise to approach it through the degree of similarity/dissimilarity in object-subject relationships than as an either/or enterprise.

13.7 Cause and Conscious Experience

Recall Chalmers' stance that materialist/ reductionist explanations are satisfactory for some phenomena but not for the hard problem. Whether one considers the question of consciousness as the cause of behavior or the question of what causes conscious experience for Whitehead, understanding cause is different from views held by traditional materialists who attribute cause to the existence of enduring material in spacetime.

> Thus, as disclosed in the fundamental essence of our experience, the togetherness of things involves some doctrine of mutual immanence… each happening is a factor in the nature of every other happening…Consider our notion of "causation." … no event can be wholly and solely the cause of another event…. But some one occasion in an important way conditions the formation of a successor.[113]

In Whitehead's philosophy, both continuity and change are given so understanding cause is not about the initiation and cessation of change or the transferring of qualities between the surfaces of two formally disconnected entities.[114] He described this by saying that events that condition the contingencies of other events are causes. Conditioning events can generally be categorized as active or passive. Active conditioning events cause alterations either by processes of a generating or transmitting type. Passive conditioning is the rest of nature relegated to background and is important because this is what did not get incorporated into the becoming event except as background of what 'is not.'[115]

Sense-objects (colors, sights, sounds, e.g.,) and perceptual objects (chairs, rocks and trees, people) may be thought of as adjectival events that describe situations and therefore reduce the contingencies of nature. By reducing the contingencies one could think of them as reducing the possible 'could be.' Scientific objects are events such as electrons that are not obvious to us through ordinary sense awareness but are the objects that physicists study. Sense-objects, for example, only qualify or describe the present but scientific objects condition future events.

Whitehead explained, "It is evident therefore that a scientific object must qualify future events. For otherwise the future contingency is unaffected by it."[116] Scientific objects condition the future because scientific objects are characterized by the

[113] Whitehead 1934/1961, p. 26.

[114] This is reminiscent of Edward Reed's 1996 regulation versus construction distinction in *Encountering the World*, Reed 1996, pp. 9–19.

[115] Tiebout 1958, pp. 46–47.

[116] Whitehead 1922, p. 32.

property of ‘ongoingness,’ similar to Pred’s onflow.[117] As described in the earlier section on Whitehead’s ontology, the suggestion that the process of quantum state evolution is a scientific object helps picture what Whitehead meant.

For Whitehead it is important that the past and the future of an actual entity’s duration are mutually exclusive events. The efficient causes of conscious experience are events that are already over by the time we are conscious of them. The events occurring prior to conscious experience must obey all of the same conditioning events that apply to nature and this includes the constant Einstein derived as the speed of light. Whitehead derived the same value differently but it is a fundamental feature of nature that events ‘take time.’ His view is that time is an abstraction and is created by events because the past and causally closed. Anytime there is a collection of contemporary actual entities, there is duration but no contemporaries have instantaneous causal efficacy.

With respect to consciousness as a cause, earlier I mentioned that Whitehead wrote of higher grades of experience that include aversions and adversions.[118] Aversions and adversions during experience form a basis and possible role for private subjective experiences of qualia in people. Since these may form the basis for the operation of preferences, of valuations of pleasantness or unpleasantness they may alter outcomes in highly individual ways. Here is the heightened kind of contrast to which Whitehead referred as essential to the height of consciousness. A thing is not just brown (discrimination) but brown and not gray, not red, or anything else it could be.

It is important to recall that Whitehead was not trying to argue that anything but the most complex entities would have the basis for the most complex versions of experience and even in people this occurred in gradients so that the height of consciousness was rare. It is in the degree of consciousness contrasts that conscious experiences may exert causal influence over future behavior. This is the final cause and (possibly) the uniqueness imparted to the subjective experience. From Whitehead’s view, efficient causes are the finished actualities, are already determined, and are in the past light cone to which a subject must conform. Outcomes are not perfectly determined, however, because causal efficacy does not stop at the ‘outside’ of the effect but enters into and incorporated as part of its internal, private production of potentia, the ‘nextness’ inherent in natural events.

The subject’s final cause is found in the relationship between the efficient cause entering into it and the subject’s private production of potentia. Whitehead saw problems with any concept of perfect accuracy or perfect precision so it is not surprising that his view implies that there is a strong, but not perfectly predictable relationship between what others dichotomize as the external world and the internal experiences of a subject. Because there is always something entirely new in each process, the outcomes of the world’s relationship with an experiencing subject are not perfectly predictable. The indefinite nature of internality and externality is part of the subject-object relationship. Whitehead stated, “Our knowledge of the body

[117] Pred 2005.

[118] Whitehead 1929a, b, p. 234 and 241.

places it as a complex unity of happenings within the larger field of nature … The body consists of the coordinating function of billions of molecules. It belongs to the structural essence of the body that, in an indefinite number of ways, it is always losing molecules and gaining molecules … the body requires the environment in order to exist. Thus, there is a unity of the body with the environment…".[119]

Scientists today, and in Whitehead's day, put a great deal of weight on material as the cause of change in states. The material outside of our bodies causes memories, perceptions, or thoughts by changing the material inside. The change in the material inside is presumed to create a representation of the outside material and therefore the outside material determines our actions, one way or another. Whitehead's philosophy makes the subject/object, body/mind, internal/external, organism/environment relationships mutually implicative. More than interacting, co-acting, or reciprocal there is a genuine and required physical continuity. This is one of the implications of Whitehead's depiction of the universe as an extensive continuum. No one aspect of these gradients can be said to have causal primacy.

Chalmers viewed the physical domain is causally closed but conscious experience might have subtle causal relevance as an internal supplement to the physical. The causal weight given material is consistent with Chalmers' faith that reductive science will successfully explain the existence of living systems and various life functions. Chalmers notes C. D. Broad's 1925 *The Mind and its Place in Nature* when framing the consciousness problem.[120] In that chapter Chalmers asserted and rejected, as would Whitehead, several possible types of materialist and dualist explanations for conscious experience. One view which Chalmers called Type-F monism, is essentially the dual-aspect information theory proposed earlier when he proposed the hard problem of consciousness.

Chalmers' suggestion that the experiences of consciousness are fundamentally related to the intrinsic properties of physical reality is similar to Whitehead's. Whitehead speculated that there is a fundamental relationship in nature between efficient cause and final cause. The efficient cause is found when objective actual entities in a subject's past light cone evoke activity in the subject's process of becoming. The efficient cause enters and becomes constitutive of an effect during the internal private production of potential outcome states.[121] In reminding us that a quark is characterized by its relationships to other physical entities, Chalmers framed the issue much as Whitehead did. If Chalmers' were to frame the concept of a quark as an event related to other events, rather than as a particle that may have relevant intrinsic properties, it would be even closer to Whitehead's natural philosophy.

---

119 Whitehead 1934/1961, p. 26.

120 Chalmers 2010, p. 133. This is the same C.D. Broad to whom Whitehead referred when he thanks Broad for the use of the term 'historical' in his 1922 *Principles of Relativity*.

121 Whitehead 1929a, p. 237.

13.8 A Few Implications of Whitehead's Natural Philosophy

Whitehead is not thought of as having had a great deal of influence over the philosophical and scientific community. Recall that Weber likened him to a seldom appearing ghost in scholarship limbo suggesting that even Whitehead experts may have placed undue emphasis on his terminology at the expense of careful attention to his key logical and ontological insights.[122] Nonetheless, interest in process approaches and Whitehead's views have increased since the late 1980s. Consistent with this observation, for example, Epperson reviewed Stapp's work on quantum mechanics and consciousness and argued that Stapp's work involved interpreting Whitehead to be consistent with the conscious-observer dependent view of reality, with which Whitehead disagreed.[123] Once so suspect Einstein, among others, rejected it, familiarity with quantum mechanical state evolution theorizing may have contributed to the rising number of psychologists and neuroscientists attempting to map Whitehead-like views onto the biology of consciousness.

For example, several authors have examined the notion that all entities may be thought of as having experience. Given Whitehead's avoidance of sharp dichotomies it is no surprise that his re-framing of experience as a term equally applicable to event particles as to people has captured the attention of authors interested in the evolution of consciousness. Sometimes under the term 'panexperientialism' authors from different scholarly perspectives have argued for the existence of gradations of conscious experience in the evolutionary tree including, for example, Griffin, Velmans, and Weber.[124]

Penrose integrated ideas from Whitehead's process view, quantum mechanical state evolution, and experimentation on microtubule activity to propose a theory of conscious experience that is consistent with the evolution of conscious experience rather than a view that it is unique to humans.[125] Finally, ecological psychologists do not mention Whitehead, yet he has made a ghost-like appearance there. For example, Reed (1996) expressed a view reminiscent of Whitehead's idea that none of the data available for understanding nature are exclusively 'in the mind' or 'out of the mind.' "An animal's actions and awareness have a rich causal substrate, not just in the animal's nervous system but in the environment surrounding the animal; however, none of these causal factors, either individually or collectively, completely causes psychological states. This is just what agency means …".[126] History has by no means given us an answer to the hard problem and Whitehead's contribution only raises more questions. Not the least of these questions is can there ever be a genuinely non-reductive explanation if one simultaneously retains any traditional notions of disconnected inert "material" as fundamental to nature? In examining Alfred

[122] Weber 2006, p. 118.

[123] Epperson 2009, p. 349.

[124] Griffin 1993; Velmans 2009; Weber 2006.

[125] Penrose 2011.

[126] Reed 1996, p.18.

North Whitehead and the history of the hard problem of consciousness, there is a sense in which we have come full circle. Whitehead's work suggests we return to contemplating a Platonic generative space with a world of becoming as being. He returns us to Aristotle and the explanatory possibilities engendered by the inclusion of both the actual and the potential, the efficient and the final cause.[127]

References

Boscovich R (1922) A theory of natural philosophy. In Petronijevíc B (ed) (trans: Child JM). Frome, Chicago (Original work written in 1763)

Broadman K (1999) Localisation in the cerebral cortex (trans: Garey L). Imperial College Press, London (Original work published 1909)

Cajal SR (1906) The structure and connexions of neurons: Nobel lecture, December 12, 1906. In: Nobel lectures, physiology or medicine 1901–1921. Elsevier Publishing Company, Amsterdam. Stable URL: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1906/cajal-lecture.pdf

Carr W, Nunn TP, Whitehead AN, Wrinch D (1922) The idealistic implication of Einstein's theory. Proc Aristot Soc New Ser 22(1921–1922):123–138, Stable URL: http://www.jstor.org/stable/4544017

Chalmers DJ (1995) Facing up to the problem of consciousness. J Conscious Stud 2(3):200–219. URL: http://cogprints.org/316/1/consciousness.html. Accessed 10 Feb 2012

Chalmers DJ (2010) The character of consciousness. Oxford University Press, New York

Dewey J, Bentley G (1949) Knowing and the known. Beacon Press, Boston

Eddington AS (1924) A comparison of Whitehead's and Einstein's formulae. Nature 113(2832):192. doi:10.1038/113192a0

Edwards L (2013) A brief conceptual history of *Einfühlung*: 18th-century Germany to post-World War II U.S. Psychology. J Hist Psychol 16(4):269–281

Einstein A (1905a) Generation and conversion of light with regard to a heuristic point of view. Ann Phys 17(6):132–148

Einstein A (1905b) Is the inertia of a body dependent on its energy content? Annalen Der Physik 18(13):639–641

Epperson M (2004) Quantum mechanics and the philosophy of Alfred North Whitehead. Fordham University Press, New York

Epperson M (2009) Quantum mechanics and relational realism. Process Stud 38(2):339–366. http://www.csus.edu/cpns/library/Epperson_QMandLogicalCausality_ProcessStudies38v2.pdf. Accessed Feb 2012

Gell-Mann M, Hartle JB (2007) Quasiclassical coarse graining and thermodynamic entropy. Phys Rev A 76:022104. doi:10.1103/PhysRevA.76.022104

Gibbons G, Will CM (2008) On the multiple deaths of Whitehead's theory of gravity. Stud Hist Philos Mod Phys 39:41–61. doi:10.1016/j.shpsb.2007.04.004

Griffin DR (1993) Founders of constructive postmodern philosophy: Peirce, James, Bergson, Whitehead, and Hartshorne. State University of New York Press, Albany

Griffiths RB, Omnès R (1999) Consistent histories and quantum measurements. Phys Today 52(8):26–31. doi:10.1063/1.882775

Heisenberg W (1927) Ueber den anschaulichen Inhalt der quantentheoretischen Kinematik and Mechanik. Zeitschrift für Physik 43:172–198. English translation in (Wheeler and Zurek, 1983), 62–84

[127] Whitehead 1929b, *The Function of Reason*.

Hoad TF (ed) (1996) “receptacle” The concise Oxford dictionary of English etymology. Oxford University Press. Oxford Reference Online. http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t27.e12530
Lashley K (1924) Studies of cerebral function in learning VI. Psychol Rev 31(5):369–375. doi:10.1037/h0070668
Lowe V (1975) A. N. Whitehead on his mathematical goals: a letter of 1912. Ann Sci 32(2): 85–101. doi:10.1080/00033797500200161
Lowe V (1985) Alfred North Whitehead: the man and his work. Johns Hopkins University Press, Baltimore
Lukerman F (1961) The concept of location in classical geography. Ann Assoc Am Geogr 51(2):194–210
Marvin EL (1927/1928) The philosophy of science lectures by Alfred North Whitehead at Harvard University (1927–28). http://whiteheadresearch.org/research/lecture-notes/index.shtml
Merzbacher E (2002) The early history of quantum tunneling. Phys Today 55(8):44–49. doi:10.1063/1.1510281
Minkowski H (1909) Space and time. Presentation to the 80th meeting of German scientists and physicians, Cologne, September 21, 1908. English translation by Lehmkuhl D, In Petkov V (ed) Minkowski spacetime: a hundred years later. Springer, New York, pp xv–xii
Newton I (1726/1999) The principia: Mathematical principles of natural philosophy: a new translation (trans: Cohen IB, Whitman A, preceded by “A guide to Newton’s Principia” by Cohen IB). University of California Press, Berkeley
Penrose R (2011) Consciousness in the universe: neuroscience, quantum space-time geometry and Orch OR theory. J Cosmol 14:1–80, Journal of Cosmology.com
Petkov V (2010) Space, time, and spacetime: physical and philosophical implications of Minkowski’s unification of space and time. Springer, New York
Plato (trans. 1871/1953) The dialogues of Plato: translated into English with analyses and introductions, vol 3, 4th edn (trans: Jowett B). Oxford University Press, Oxford
Pred RJ (2005) Onflow: dynamics of consciousness and experience. Minnesota Institute of Technology Press, Cambridge, MA
Reed E (1996) Encountering the World. Oxford University Press, New York
Reinhardt M, Rosenblum A (1974) Whitehead contra Einstein. Phys Lett 48A(2):115–116. doi:10.1016/0375-601(74)90425-3
Sahraiea A, Hibbard P, Trevethan C, Ritchiea K, Weiskrantz L (2011) Consciousness of the first order in blindsight. Proc Natl Acad Sci 107(49):21217–21222. doi:10.1073/pnas.1015652107
Schmid M, Mrowka S, Turchi J, Saunders R, Wilke M, Peters A, Ye F, Leopold D (2010) Blindsight depends on the lateral geniculate nucleus. Nature 466:373–377. doi:10.1038/nature09179
Schrödinger E (1926) An undulatory theory of the mechanics of atoms and molecules. Phys Rev 28(6):1049–1070
Sherrington CS (1925) Remarks on some aspects of reflex inhibition. Proc R Soc Lond B Biol Sci 97(686):519–545, Stable URL: http://www.jstor.org/stable/81057
Snyder G (1990) Burrhus Frederic Skinner: the man behind the science. Performance Management Magazine 89(2):19–25, http://pmezine.com/archive/performance-management-magazine-springsummer1990-vol8-no2/index.html
Stapp HP (2005) Quantum interactive dualism: an alternative to materialism. J Conscious Stud 12(11):43–58. doi:10.1111/j.1467-9744.2005.00762.x
Tiebout HM (1958) Appearance and causality in Whitehead’s early writings. Philos Phenomenol Res 19(1):43–52
Velmans M (2009) Consciousness, brain and the physical world. Philos Psychol 3:77–99. doi:10.1080/09515089008572990
Weber M (2006) Whitehead’s onto-epistemology of perception and its significance for consciousness studies. New Ideas Psychol 24:117–132
Weber M, Weekes A (2009) Process approaches to consciousness in psychology, neuroscience, and philosophy of mind. State University of New York Press, Albany

Whitehead AN (1889a) On the motion of viscous incompressible fluids: a method of approximation. Q J Pure Appl Math 23:78–93
Whitehead AN (1889b) Second approximation to viscous fluid motion: a sphere moving steadily in a straight line. Q J Pure Appl Math 23:143–152
Whitehead AN (1906) On mathematical concepts of the material world. Royal Society of London, Dulan, London
Whitehead AN (1916) Presidential address. Section A of the British Association for the Advancement of Science. http://www.biodiversitylibrary.org/item/96027, http://www-history.mcs.st-andrews.ac.uk/Extras/BA_1916_2.html
Whitehead AN (1920/1955) The concept of nature. Cambridge University Press, New York
Whitehead AN (1922) The principle of relativity with applications to physical science. Cambridge University Press, London
Whitehead AN (1925) Science in the modern world. The MacMillan Company, New York
Whitehead AN (1929a) Process and reality: an essay in cosmology. Cambridge University Press, Cambridge, MA. Corrected Edition. Griffin DR, Sherburne D (eds) Free Press, New York, 1978.
Whitehead AN (1929b) The function of reason. Princeton University Press, Princeton
Whitehead AN (1931) Objects and subjects. Proc Address Am Philos Assoc 5:130–146, Stable URL: http://www.jstor.org/stable/1483082
Whitehead AN (1934/1961) Nature in life. In: Anshen RN (ed) Alfred North Whitehead: his reflections on man and nature. Cambridge University Press, New York, pp 10–45
Whitehead AN (1936) Letter to his former student, Henry Leonard. Leonard sent Whitehead a copy of his recent essay on logical positivism. Archival Manuscript Material (Collection) Unprocessed. MSS85573; Repository: Library of Congress Manuscript Division, Washington, DC. http://hdl.loc.gov/loc.mss/mss.home
Whitehead AN (1948) Einstein's theory. In: Science and philosophy. Philosophical Library Inc, New York, pp 303–314
Whitehead TN (1963) Now I am an American: habit and change. Harvard Business School, Cambridge, MA
Whitehead AN, Russell B (1910, 1912, 1913) Principia mathematica, 3 vols. Cambridge University Press, Cambridge. Second edition, 1925, vol 1; 1927, vols 2, 3. Abridged as Principia mathematica. Cambridge University Press, Cambridge, 1962
Whitehead AN, Lodge O, Nicholson JW, Head, Stephen A, Carr HW (1919) Symposium: time, space, and material: are they, and if so in what sense, the ultimate data of science? Proceedings of the Aristotelian society, supplementary volumes, 2, 4:108. Stable URL: http://www.jstor.org/stable/4106442
Zammito JH (1992) The genesis of Kant's critique of judgment. Cambridge University Press, Cambridge, MA
Zurek W (2002) Decoherence and the transition from quantum to classical—revisited. Los Alamos Sci 27:2–25. http://arxiv.org/ftp/quant-ph/papers/0306/0306072.pdf. Accessed 23 Mar 2012

Chapter 14
The 'Hard Problem' and the Cartesian Strand in British Neurophysiology: Huxley, Foster, Sherrington, Eccles

C.U.M. Smith

'But strictly, we have to regard the relation of brain to mind as still not merely unsolved but still devoid of a basis for its very beginning.'

Charles Sherrington 1933, *The Rede Lecture*, p. 32

It began, like so much else, in the seventeenth century. Both Galileo and Descartes, although the focus of much ecclesiastic wrath, were, as Gaukroger points out, in fact trying to save the Church from itself.[1] Galileo states explicitly that his endeavour was only to correlate certain 'accidents' and that all else must be left to 'higher science' than his, that is theology.[2] Descartes, similarly, hoped his natural philosophy would replace the time-worn and outdated philosophy of the schools.[3] When he became aware of Galileo's condemnation by the Holy Office he was, as he wrote to Mersenne, devastated and was strongly inclined to commit his writing to the flames. 'I would not' he writes, 'publish anything contradictory to the teachings of the Church'. He did not, of course, entirely keep his promise. He took good care to reside in obscure and remote Dutch seaside resorts and kept his most revolutionary work, *L'Homme*, back to be published posthumously 12 years after he met his end in the freezing Swedish winter.

In the century succeeding his death his work had very little influence on the biological sciences. It seemed, as Niel Stensen wrote, to be no more than a dream: 'that Philosopher [Descartes] hath rather devised, in his *Treatise of Man*, such an engine

Author was deceased at the time of publication.

[1] Gaukroger 1995.

[2] Galilei 1638, Day 3, Cor. 3.

[3] Gaukroger 1995.

C.U.M. Smith (✉)
Emeritus, Department of Vision Sciences, Aston University,
Aston Triangle, B4 7ET Birmingham, UK
e-mail: a.m.jennysmith@gmail.com

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_14,

that performs all the actions Men are capable of than described Man as he really is.'[4] In Descartes' defence it is not unreasonable to remark that this is no more than he intended. Stensen is also scathing about Descartes' treatment of the pineal, writing that the anatomy is entirely imaginary and that it could not play the psychophysiological rôle that Descartes had devised for it. It was not until the nineteenth century that Descartes' neurophysiology at last found a powerful champion.

This champion was Thomas Henry Huxley (1825–1895). As we noted in Chap. 8, Huxley praises Descartes 'as a great and original physiologist', as one who had done for 'the physiology of motion and sensation that which Harvey had done for the circulation of the blood and opened up that road to the mechanical theory of these processes which has been followed by all his successors'.[5] Huxley occupied a powerful position in the British scientific establishment. At the top of the building in South Kensington which was to become Imperial College he trained the teachers who were to go out into the schools and colleges to renew British biology.[6] Among the many subsequently distinguished people who worked with Huxley as demonstrators was Michael Foster. Foster succeeded Huxley as Fullerian professor of physiology at the Royal Institution in 1869, and in 1870 Huxley recommended him for the post of praelector in physiology at Trinity College, Cambridge. There he revolutionized physiological education along Huxleyan lines, ensuring that all students received thoroughgoing practical tuition in the laboratory. In addition to writing his highly influential *Textbook of Physiology* (1876), founding the *Journal of Physiology* (1878), and jointly editing Huxley's *Scientific Memoirs* (1898–1902) Foster, like his mentor, developed a deep interest in the history of his subject.

This interest came to fruition when Foster was invited to give the Lane lectures in San Francisco in the autumn of 1900. These were published in 1901 as *Lectures on the History of Physiology during the Sixteenth, Seventeenth and Eighteenth Centuries*.[7] 'What we know and what we think,' he writes at the beginning of this volume, 'is not a new fountain gushing fresh from the barren rock at the stroke of the rod of our own intellect, it is a stream which flows by and through us, fed by the far off rivulets of long ago'.[8] He follows up this very Huxleyan sentiment later in his book with another Huxleyan assessment, when in his chapter on the history of neurophysiology he writes, '… very little change in the details of Descartes' exposition and * some of that hardly more than a change in the terminology would convert that exposition into a statement of modern views'… Descartes' exposition will not appear so wholly different from the one which we give today.'[9]

[4] Stensen 1669; English translation may be found in Gotfredsen 1950.

[5] Huxley 1874, p. 201.

[6] See account in Desmond 1997. And not only teachers: H.G. Wells was one of Huxley's last students at South Kensington and remarks that it 'was beyond all question, the most educational year of my life' (Desmond 1997, p. 158).

[7] Foster 1901/1970.

[8] Foster 1901/1970, p. 1.

[9] Foster, ibid., p. 298.

Michael Foster's physiology classes were attended by many students who later left their mark on the subject. None, however, were to surpass Charles Scott Sherrington (1857–1952). After graduating with first class honours in Natural Sciences from Michael Foster's classes he left to take medical degrees at St Thomas Hospital in London. After a number of years (1887–1895) as lecturer in physiology at St Thomas's and Superintendent of the Brown Institute in London[10] he was appointed Holt Professor Physiology at Liverpool (1895–1913). In 1913 the Waynflete Chair of Physiology at Oxford became vacant and Sherrington was elected by a unanimous decision. He remained at Oxford until he retired in 1936. During his retirement he kept up a wide correspondence and continued working on his lifetime interests in philosophy and literature.[11] He died at the age of 95 in 1952.

Charles Scott Sherrington could almost be taken as the type example of a 'man of many parts'. Over the years of his professional appointments at St Thomas's and the Brown Institute and at the Universities of Liverpool and Oxford he produced with his colleagues over 300 incisive papers on neuromuscular physiology, culminating in his great work of synthesis, *The Integrative Action of the Nervous System*[12] and the award of the Nobel Prize in Physiology or Medicine in 1932.[13] Many of these works are collected in Denny-Brown's volume '*The Selected Writings of Sir Charles Sherrington*'[14] Denny-Brown quotes with approbation the words of Professor Camis who, writing of the vestibular system, observes that '… over a considerable period of time, the work and ideas of one man [Sherrington] entered into the whole structure of the physiology of the nervous system …'[15] His students, moreover, included many who subsequently became world-leaders in their subjects, including such famous names as those of Wilder Penfield, Granit, Liddell and, of course, Eccles, whose long-term confrontation with the 'hard problem' we will consider in some detail in the last part of this chapter.

But in addition to his Nobel prize-winning contributions to neuromuscular physiology Sherrington was also a considerable poet and philosopher and historian of science. His poetry came early and was published in a small volume, *The Assaying of Brabantius*, in 1925 (2nd edition in 1940).[16] The early verses look back in their sentiment to the late Victorian and Edwardian eras and their language to a yet earlier period. As Fuller points out there are echoes of both Coleridge and Keats[17] and nothing of the anguished disillusionment of the war poets or the fractured barrenness of

[10] The major activity of the Brown Institute was veterinary medicine. It was established in 1871 and directed by a number of distinguished figures including Sir John Burdon-Sanderson, Victor Horsley and Charles Sherrington.

[11] Further detail is given in Nobelprize.org

[12] Sherrington 1906; 2nd edition 1947.

[13] Sherrington, Nobel Prize.

[14] Denny-Brown 1939.

[15] Ibid., p. ix.

[16] Sherrington 1925.

[17] Fuller 2007.

Eliot's 1922 *Waste Land*. Sherrington writes of Oxford bells waiting for fresh exploits to celebrate[18] and of how the call of duty justifies the sacrifice of young life – very different from the bitter anger of Wilfred Owen's *Dulce et decorum est*!

Some of this almost Tennysonian language and sentiment informs his last great philosophical statement published in 1940 at the start of another war. A second edition came in 1951. This volume, somewhat ambiguously titled *Man on his Nature*, originated as a series of Gifford Lectures delivered at the University of Edinburgh in 1937/8.[19] In it he sums up, in his 80th year, a life-time of thought about man's nature and his place in Nature and, in particular, about the 'hard problem' of the relation of mind to brain.

In addition to the mannered sub-Tennysonian prose, Sherrington makes use of his research into the great sixteenth-century French physician, Jean Fernel, to give historical depth and perspective to his thesis.[20] He ultimately published this research as a scholarly monograph, *The Endeavour of Jean Fernel*, in 1946.[21] The research represented in this book, which also includes interesting passages on the seventeenth century and René Descartes, enabled Sherrington to situate mid-twentieth century neurophysiology in the 'longue durée' of scientific history. For, as Frank Kermode points out, '… history continues to be the means by which we recognize what is new from what is not.'[22]

One of the most significant 'newnesses' that Sherrington stresses throughout *Man on his Nature* is the modern distinction between 'life' and 'mind'. With Fernel in the sixteenth century this distinction, writes Sherrington, did not exist. He would, he says, have smiled at our problem: 'For him there is no difference between thought and the rest of living.'[23] How different things are in our twenty-first century! Indeed, how different they were in Sherrington's mid-twentieth century. The deep and powerful river of chemical, biochemical and molecular biological research, gathering pace throughout the last two centuries, has shown, without the shadow of a doubt, that life is a complicated chemistry, and that this chemistry, as Sherrington again and again emphasizes, is at one with and, in its essence, no different from, the simpler chemical processes which are ceaselessly at work in the Earth's crust.[24]

[18] In a poem written in 1916, 'Now in the cloister few feet that roam,' he writes of the Oxford bells hearing 'from far the filial bugles blow'. But by the war's end in 1917 and 1918, his poems 'Dawn's Red' and 'I met a man by yonder mill', the language, although still sub-Tennysonian, tells of a much darker mood.

[19] Sherrington 1951.

[20] This historical approach to the abiding issues of philosophy would have pleased Thomas Huxley who averred in one of his essays that 'that there is assuredly no more effective way of clearing one's own mind on a subject than by talking it over, so to speak, with men of power and grasp, who have considered it from a totally different point of view … the parallax of time helps us to the true position of a conception as the parallax of space helps us to that of a star.' (Huxley 1874, p. 202).

[21] Sherrington 1946.

[22] Kermode 1971, p. 64.

[23] Sherrington 1951, p. 240.

[24] Sherrington 1951, chapter 5.

Underlying this ceaseless activity, this 'coming to be' and 'passing away', and ultimately responsive for it, is, says Sherrington, the phenomenon we know as energy. It underlies, as he phrases it, 'Earth's reshuffling'.[25] Energy, he argues, is responsible for all the activity we see in the world of Nature, from the movement of the galaxies to the fall of a raindrop, from the chemical reactions ceaselessly churning the inorganic world to those equally ceaselessly occurring in living organisms, from the life and reproduction of the microbe to the activities of our sleeping and, far more, in our waking brains. It is also responsible, he says, for the great 'blind' process of organic evolution. Sherrington fully accepts the neo-Darwinian thesis that the living world we know today has evolved by chance and happenstance, by 'random variation and selective retention'. Throughout that world he sees, moreover, an overpowering '"urge" or "zest-to-live"'.[26] His view of the living world is bleak: he sees only a pitiless 'war of all against all'. This is detailed in many parts of *Man on his Nature* but never more starkly than in the last chapter where he reviews the contest between what he describes as 'an amoeboid speck' and humanity. He writes, of course, of the contest between the 'million-murdering' *Plasmodium malariae* and almost one quarter of the human population.[27] No quarter is given in the struggle for existence, no distinction is made between so-called 'higher' and 'lower' forms of life.

But, says Sherrington, there is one exception to this energy-based analysis of the world – mind. Not only in *Man on his Nature* but also his celebrated *Rede Lecture* Sherrington is adamant that no investigation of the physiology of the brain, no matter how subtle, will ever discover the whisper of a thought or a feeling.[28] In *Man on Nature* he writes, as he did in his *Rede Lecture*, that 'the mind is something with such manifold variety, such fleeting changes, such countless nuances, such wealth of combinations, such heights and depths of mood, such sweeps of passion, such vistas of imagination, that the bald submission of some electrical potentials recognizable in nerve-centres as correlative to all of these may seem to the special student of mind almost derisory.'[29] A further sixty or so years of research has revealed so much more of the huge, organised complexity of the human brain, down to the molecular level, that we may not be as convinced, as Sherrington seemingly was, that there is an unbridgeable gap between 'the manifold variety' of our subjective lives and the intricate physical-chemical wave patterns of our living grey matters.

---

[25] Ibid., chapter 5. It might be said that a more fundamental principle is encapsulated in thermodynamics second law: entropy inexorably increases.

[26] Sherrington hesitates to use the more Nietszchean expression 'will-to-live' because it may be mistaken for implying some 'conscious' intention.

[27] Modern medicine and environmental control has reduced this atrocious sum of human death and suffering somewhat since Sherrington wrote in the first half of the twentieth century. A recent summary in *Nature* shows that about a million malaria-induced deaths still occur each year (Shetty 2012).

[28] Sherrington 1933.

[29] Sherrington 1951, p. 238.

Nevertheless, even if the fretted physical activity of our brains may, in future years, be plausibly correlated with the patterns of our subjectivities, one glaring problem, as Sherrington insists, remains. 'The mental' he writes, is just 'not examinable as a form of energy. That in brief', he goes on, 'is the gap which parts psychiatry and physiology. No mere running round the cycle of 'forms of energy' takes us across the chasm'.[30] This is the nub of Sherrington's problem, as it is still the nub of the 'hard problem' today. 'Thoughts, feelings and so on are not amenable to the energy (matter) concept. They lie outside it. Therefore they lie outside Natural Science.'[31] This, as Sherrington, goes on to say, 'is embarrassing for biology'. It is clear that we and our fellow human beings, along with the higher mammals and perhaps some of the birds, are blessed (or cursed) with mind – that is cognition and sensation. Since Darwin, as Sherrington was the first to acknowledge, biologists acknowledge that an unbroken line of organic forms leads back through the aeons of geological time to an origin in the inorganic surface chemistry of the cooling planet. 'When', to use Darwin's phrase, 'did consciousness commence'?

Sherrington has no real answer to this conundrum. He returns to it time after time. 'There is', he writes in his *Rede Lecture,* 'so far as I know, in the chemical, physical properties, or in the microscopical structure, no hint of any fundamental difference between the non-mental and the mental regions of the brain.'[32] Three quarters of a century later, with a huge accession of knowledge of the brain's structure and functioning down to the molecular level and beyond, brain scientists would still hasten to agree. Sherrington sums up, putting himself forward, somewhat absurdly, as the 'plain main': 'I notice with the relief of a plain man' he writes, 'that often busy common sense, naïve and shrewd, has the world with him; the lawyer, the doctor, the economist, and indeed the man in the street generally… [accept] that brain and mind go other.' But 'that acceptance', he continues 'nowise removes the enigma of how it is they do so.'[33]

Sherrington is, of course, anything but a 'plain man' or 'man in the street'. He is the greatest expert in muscular movement of his generation. Hence when he writes of moving a limb, we listen.[34] He is supremely well informed about the myriad physiological events which bring this movement about. Yet, as he says, subjectively we know nothing of these neurophysiological events. He, like the rest of us in this situation, is merely aware of where the limb is in space, of our intention to move it and of its consequent movement. He remarks, from his historian's perspective, how 'the analysis of the perceptible world seems to have outstripped his analysis of his own mind'[35] and while the analysis and understanding of the neuromuscular physiology has advanced by leaps and bounds, 'ways of thought regarding the mind

[30] Sherrington 1951, p. 238.

[31] Ibid., p. 239.

[32] Sherrington 1933, p. 28.

[33] Ibid., p. 25.

[34] Sherrington, 151, pp. 253–4.

[35] Ibid., p. 255.

would seem in the last 2000 and odd years not to have changed to anything like the same extent.'[36] Sherrington ends this part of his argument with a strong statement of the classical Cartesian position:

> The sun's energy is part of the closed-energy cycle. What leverage does it have on mind? Yet through my retina and brain it is able to act on my mind. Conversely my thinking 'self' thinks that it can bend my arm. Physics tells me that my arm cannot be bent without disturbing the sun. Physics tells us that unless my mind is energy it cannot disturb the sun. My mind then does not bend my arm. If it does, the theoretically impossible happens. Let me prefer to think that the theoretically impossible does happen. Despite the theoretical I take it that mind does bend my arm, and that it disturbs the sun.[37]

Sherrington has completed one part of his argument and arrived at what seems to be an ineradicable duality. He concludes that we just have to accept the mystery. In the second (1947) edition of *The Integrative Activity of the Nervous System* he remarks, 'That our being should consist of two fundamental elements offers, I suppose, no greater inherent improbability than that it rest on one only.'[38] In *Man on his Nature* he merely observes that Nature has simply 'evolved us as compounds of energy and mind'[39] so that 'our world, which in our experience is one world, [is] a diune world, a world of outlook and inlook'[40] But Sherrington has not completed his argument quite yet. Although, as we have seen, a convinced neo-Darwinian, he nevertheless believes that he can detect a beckoning hope of an ultimate purpose for mind on Earth.

We are evolved as amalgams of mind and matter and no one has expressed the suffering and futility of the living world, its individual 'zests' for life pitted against others better than Sherrington in *Man on his Nature*. Yet he seems to see the glimmerings of a resolution. Out of the immense struggle for existence in an uncaring biosphere where, as his example of that million-murdering speck of matter, *Plasmodium malariae*, makes clear, the notions of higher and lower forms of life have no meaning, it may just be, he thinks, that the evolution of the human mind signals a break through to a new order. He notes, first, that 'predatory forms of life', he instances the hawk, the eagle, the leopard and the tiger, do not live community lives. Humans do. Humans share values. A solitary human, as Aristotle long ago remarked, can have no concept of morality. Humans are only the forms on the planet's surface whose 'zest to live' includes a zest to see other lives succeed and not human lives only. Humanity is 'slowly drawing from life the inference that altruism, charity, is a duty incumbent on thinking life'.[41] From unthinking nature,

---

[36] Ibid., p. 255.

[37] Ibid., p. 258.

[38] Some have suggested that a way out of Sherrington's impasse is to use metaphors derived from the physicists' familiar recourse to higher dimensions than our commonsensical four. The physicist, of course, is merely generalising from the Cartesian method of locating a dimensionless point in four-dimensional space-time.

[39] Ibid., p. 261.

[40] Ibid., p. 302.

[41] Ibid., p. 288.

free of values, free of assessments of what is 'higher' and what is 'lower', all these attributes have emerged. Clearly, we are here in the vicinity of what has become known as the weak anthropic principle.[42] We are also in the vicinity of modern theories of how altruistic behavior can emerge in a biosphere created by the contest of 'selfish genes'. We must, however, leave these twenty-first century arguments here for they are outside the remit of this chapter. But we can see the direction of Sherrington's thought. His evolutionary neurobiology hints that the appearance of mind in the world, at the human level, may signal the beginnings of a transcendence of Nature's 'war of all against all'. The individual's 'zest-to-live' transposed on the level of consciousness perceives a 'loftier responsibility', a responsibility which, as Sherrington writes in the closing passage of his great book, '… we cannot devolve, no, not as once was thought, even upon the stars'.[43]

Man on his Nature sums up Sherrington's views after a long life devoted to both science and the humanities. He offers no solution to the age-old conundrum presented by the 'hard problem'. Indeed, he points out that with the huge growth of chemical science since the times of his hero, Jean Fernel, and, even more, with the establishment of neo-Darwinian evolution as the foundation theory in the biological sciences, the problem has grown more acute. He recognizes the full depth of Man's loneliness in 'a world he never made' and from both literary and scientific standpoints he understands the strange and, perhaps, ultimately lethal contradiction of a being evolved by the neo-Darwinian process turning its back, so to speak, on that process and hoping to survive in its despite. The pursuit of altruism, Sherrington's final hope for the future of humanity (and in a sense, he would argue, of the future of the planet) involves, as he writes, 'an adverse criticism of [man's] one process of creation. He is becoming an adverse critic of his own "zest-to-live". But a life without "zest-to-live" will undoubtedly perish'.[44] The evolution of consciousness has brought both the contradiction and the means for its transcendence.

Sherrington's wide-ranging views on neuroscience, history, literature and philosophy must have formed the substance of the lively Saturday afternoon meetings which he and his wife held for students during his time as Waynflete professor at Oxford. One of the most brilliant of these students was the young John Carew Eccles, newly arrived from Australia. Eccles is the last member of the sequence master-pupil to be discussed in this chapter. It is thought Sherrington would have liked him to follow him in the Waynflete Chair when he retired in 1936. But it was not to be.

Eccles was born in Melbourne on January 27, 1903. Gaining first class honours MB BSc at Melbourne University at the early age of 22, he was immediately awarded a Victorian Rhodes scholarship to Oxford. After two further years of formal tuition he once again achieved first class honours, this time in Natural Science. Taking up a junior research fellowship at Exeter College he became Sherrington's

[42] See Barrow and Tipler 1988.

[43] Ibid., p. 305.

[44] Ibid., p. 294.

last research student from 1928–1931.[45] During this period he wrote eight research papers with Sherrington and submitted his D.Phil. thesis on excitation and inhibition in the central nervous system. Although, as mentioned above, Sherrington is thought to have favoured Eccles' election as his successor in the Waynflete Chair, this did not happen, and in 1937 Eccles returned to Australia where he took the position of Director of the *Kanematsu Memorial Institute of Pathology* in Sydney. From there he went to New Zealand as Professor of Physiology at the University of Otago (1944–1951) and from there to Canberra at the Australian National University (ANU) as professor of Physiology (1951–1966). In 1966, finding that it was impossible to elude ANU's compulsory retirement age of 65, and feeling it was far too early in life to spend his time cultivating an allotment, he resigned and moved to the Institute of Biomedicine Research in Chicago and from there, in 1968, to the state University of New York (SUNY) in Buffalo. When he finally retired from SUNY in 1975 at the age of 72 he migrated to Contra in Switzerland where he remained active, as we shall see below, until he died on May 2, 1997, at the age of 94.[46]

Eccles is best-known for his researches into the microphysiology of synapses. It was for this work that, in 1963, he shared a Nobel Prize in Physiology or Medicine with the biophysicists Alan Hodgkin and Andrew Huxley. But, like his master, Charles Sherrington, he was always more than a laboratory scientist. In a 1977 summing up of his scientific life he remarks (possibly with the benefit of hindsight) that 'the field which lured me into neurophysiology fifty years ago was the mind-brain problem'[47] and, as early as 1951, he had published a paper in *Nature* arguing for a dualist interpretation.[48] His first book, derived from the 1952 Waynflete lectures in Oxford, bearing the revolutionary title, *The Neurophysiological Basis of Mind*, states in the Preface his consuming interest: 'the most fundamental questions that man can ask: What manner of being are we? Are we really compounded of 'substances', spirit and matter?'[49] We can see that, although Eccles may not have had the extraordinary of depth of culture of his great mentor, his crucial interest was very much the same. Indeed, some 9 days before Sherrington died, Eccles had visited him, finding him mentally as sharp as ever, and recorded that he had been strongly urged to continue his (Sherrington's) endeavours in the philosophy of mind.[50]

Eccles's laboratory research, like that of Sherrington, largely concerned the neurophysiology of muscular control.[51] In was natural, therefore, that the great and puzzling problem of the conscious control of the muscular system, of the control of behavioural movement, was foremost in his mind. How was this possible, how was

[45] Further biographical detail may be found Nobel Foundation http://nobel.sdsc.edu/, in Anderson and Lundberg 1997; Borck 1999.

[46] Further details of Eccles' life and personality may be found in Fillenz 2012.

[47] Eccles 1977.

[48] Eccles 1951.

[49] Eccles 1953, p. vi.

[50] Eccles and Gibson 1979, p. 183.

[51] Eccles scientific papers.

freedom of action possible, in a materialistic universe? It was the same problem which had puzzled (and defeated) his predecessors, Sherrington, Huxley, Descartes and many others. The philosopher Ted Honderich expresses the problem well when he writes of ‘mental indispensability’.[52] Is there not, at some level, a contradiction in believing that my writing these words is merely an automatism? Am I not in some sense responsible for my own actions? Is there at some depth a contradiction in someone believing that he is a zombie? This seems but common sense to most of us and so it did to Eccles. Yet how could this common sense be squared with the findings of the physiological laboratory? Did not the whole school of Sherringtonian neurophysiology build on the reflexology of the nineteenth century, on the tireless experiments of Marshall Hall and their application to the cerebrum by Laycock, Sechenov and Pavlov? Did not the more advanced techniques of twentieth-century neurohistology and microphysiology, of which Eccles himself was an acknowledged world-expert, show that in the passage from sensory input to motor output there was no mysterious gap into which a non-material ‘Will’ might insert itself? The neuromuscular system was materialistic, through and through, and as such must obey the laws of the physical world. We, like all other parts of the material world, must obey the iron rules of physical determinism. This, as Kant remarked, seems very like a nightmare and one, moreover, from which there is no awakening. Thomas Huxley had very much the same opinion and he, too, could see no way out. How could ‘mind’ influence brain, and through it a world, bound by the conservation laws of matter and energy?

For many years Eccles could see no way out of the dilemma. How could his philosophical and theological convictions (Eccles was from the beginning a faithful member of the Roman Catholic Church) be made consistent with his neurophysiology? Then, in the 1980s, he began to see the possibility that the application of quantum mechanics to the brain might point to a resolution.[53] He was, according to his later collaborator Friedrich Beck, inspired by reading Henry Margenau’s 1984 book ‘*The Miracle of Existence*’.[54] Here, for the first time, he seemed to glimpse a way of escaping the iron laws of physical determinism. For Margenau, in addition to being a philosopher and educationist, was also a top-rank quantum physicist. Accordingly when, in 1991, Eccles met Beck, another sympathetic quantum physicist, he began a collaboration in an attempt to determine whether the fundamental break with classical matter-theory made by physicists in the early part of the twentieth century might point the way forward.[55] In particular, he wondered whether the Copenhagen interpretation of quantum mechanics whereby the conscious mind was somehow involved in bringing about this ‘collapse of the wave-function’[56] might provide the break-through for which he had searched so long. Here, it seemed, at

---

[52] Honderich 1988, p. 91 etc.

[53] See also Chap. 15, Sect. 4.4, this volume.

[54] Margenau 1984.

[55] Beck 2000.

[56] See Chap. 18.

long last was a scientifically respectable instance of the direct effect of consciousness on a physical system. Could it not be that in some way the same effect was at work in the recesses of the brain?

The Copenhagen interpretation of the significance of conscious observation attempts to account for one of the deepest features of the quantum world: Heisenberg's uncertainty principle. Put simply, this states that it is impossible to determine both the spatial position and the momentum of a fundamental 'particle' at one and the same time. If one is nailed down, the other escapes, and vice versa.[57] Observation, however, nails down the position; the wave function collapses to a point particle, but the momentum of that particle is then unknowable. Before the observation is made, the position of the particle is, however, totally unpredictable. Classical causality does not apply. Heisenberg uncertainty, however, only becomes significant at the ultramicroscopic level, at the level of fundamental 'particles' such as the electron. At more macroscopic levels the uncertainty becomes too minute to be detected. Nevertheless, Heisenberg uncertainty lies at the root of our world.

However, although the uncertainty diminishes almost to nothing in the human-sized world, or in the world of conventional anatomy, physiology, histology and cytology, Eccles began to wonder whether it might not be significant at the molecular level and, in particular, at the level of the synapse's microstructure. Could quantum indeterminacy at this level provide a solution to his long-life question? Could consciousness in some way affect ultramicroscopic neurophysiology and, multiplied up through all the gyres of cerebral structure, influence neurophysiological outcomes?

But what ultramicroscopical structures could be sensitive to quantum affects? At first Eccles suggested that quantum fluctuations could affect synaptic vesicles.[58] But the numbers were against him: quantum affects would be many orders of magnitude smaller than the diffusional and thermal forces acting on these structures. Eccles then turned his attention to what he describes as the 'exquisite paracrystalline design' of the presynaptic grid.[59] Making use of his knowledge of the myotactic reflex (on which he was one of the world-experts) he observed that the activation of each synaptic bouton caused the release of at most a single 'quantum' of neurotransmitter and that the probability of release varied between 0.005 and 0.5[60] Eccles asked how this could be. Many tens or perhaps hundreds of synaptic vesicles are held in the presynaptic grid. There must, Eccles speculated, be 'some subtle functional organization … controlling the exocytosis of the embedded vesicles.'[61] In his 1992

[57] Heisenberg uncertainty is not due to the imprecision of our observational instruments; it is the way things are in the world.

[58] Eccles 1986.

[59] Eccles 1994, p. 146.

[60] See Jack et al. 1981. A quantum of neurotransmitter consists of from 5 000 to 10 000 transmitter molecules and is contained in a single synaptic vesicle. It is released as a unit into the synaptic gap.

[61] Eccles 1989.

paper with the physicist, Friedrich Beck, Eccles writes that 'the preparation for exocytosis means bringing the paracrystalline PGV (presynaptic vesicular grid) into a metastable state from which exocytosis can occur'.[62] And, to cut a long story short, it is this 'metastable state' that Eccles believed to be subject to quantum uncertainty.[63] 'It is not proposed' he writes, 'that the mental events initiate activity at a synapse by excitatory action either on the pre-synaptic or post-synaptic elements of a synapse … [but] that the mental events merely alter the probability of a vesicular emission that is triggered by a pre-synaptic impulse.'[64]

In his final papers Eccles attempted to develop his neuroquantology further. Having established to his own satisfaction how quantum level phenenoma could influence action at the synapse and thus open a gap in the iron laws of physical determinism for the mind to exert its influence, he wanted to show how this latter influence could be multiplied up to have a larger scale effect. To this end he made use of the finding by Fleischhauer and others[65] that the apical dendrites of cortical pyramidal cells in layer V are clumped together to form cylindrical aggregates known as dendritic bundles or, in Eccles' terminology, 'dendrons'. These bundles, which are clearly visible in vertical and tangential sections of the cortex, consist of three to nine apical dendrites in layer IV and more in the upper layers. Eccles proposes that it is not just one bouton that is affected by what he calls 'the quantum probability field' but all the many thousands of boutons on an apical dendrite and, more extensively still, all the many tens of thousands on a dendron (Fig. 14.1). He thus proposes that it is in this way that the mind interacts with the brain without breaking the conservation laws of physics. This interaction is 'amplified', he concludes, by 'conventional microcircuitry' to engender behavioural activity.

In his final papers Eccles introduces the concept of a 'psychon'.[66] This, he writes, is a 'unitary mental event' correlated with the action of a single dendron.[67] Psychons represent the innumerable mental experiences we live through. Eccles calculates that there are 40 million dendrons in the human cerebral cortex and there are, consequently, the same number of psychons. Our subjectivity is composed of a shifting mosaic of these psychons in various permutations and combinations. Where have we heard ideas like these before? Did not Herbert Spencer (Chap. 10) put forward rather similar ideas (without the sophisticated neuroscience or quantum physics) in the late nineteenth century?[68]

---

[62] Beck and Eccles 1992, p. 11358.

[63] Beck 1996, discusses these quantum possibilities in some detail. A review may also be found in Smith 2009.

[64] Eccles 1994, p. 73.

[65] Fleischhauer 1978; Feldman 1984.

[66] Eccles 1990, 1994, p. 87.

[67] The term 'psychon' had, in fact, been coined, unbeknownst to Eccles, by Mario Bunge in his 1980 book on the mind-brain problem (Bunge 1980).

[68] See Smith 1982.

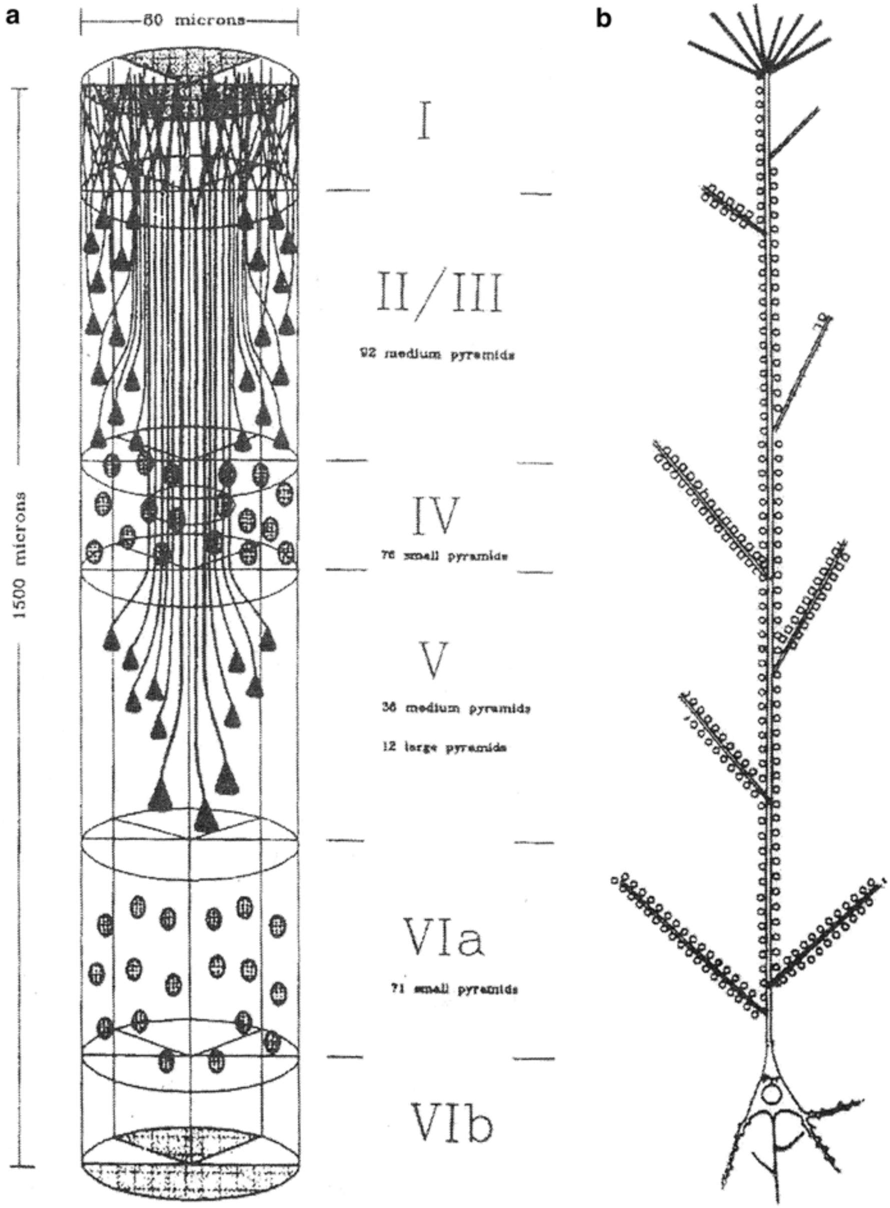


Fig. 14.1 Eccles' diagrammatic representation of a 'dendron'. The six laminae of the cerebral cortex are labelled on the *right-hand side* of the figure. The apical dendrites of the pyramidal cells in laminae II/III and V bunch together to form a bundle or 'dendron'. (**b**) Eccles' diagrammatic representation of a lamina V pyramidal cell. The apical dendrite and its branches are studded with spines onto which synapses are made. The central shaft of the dendrite ends in a 'tuft'. The intermeshing tufts are also indicated in (**a**). (**c**) On the *left* is a horizontal section of rabbit parietal cortex showing the bundles of dendrites as *white* vacuities (where the dendrite has dropped out of the section). The micrograph on the *right* is a vertical section through the same cortex showing a bundle of dendrites. Both micrographs at a magnification of 350×. (**d**) Electron-micrograph of an apical dendrite in rat visual cortex showing a spine (*S*) with a terminal synapse. The submicroscopic complexity of the cortex is well shown and this anatomical complexity is continued down to the molecular level and beyond.

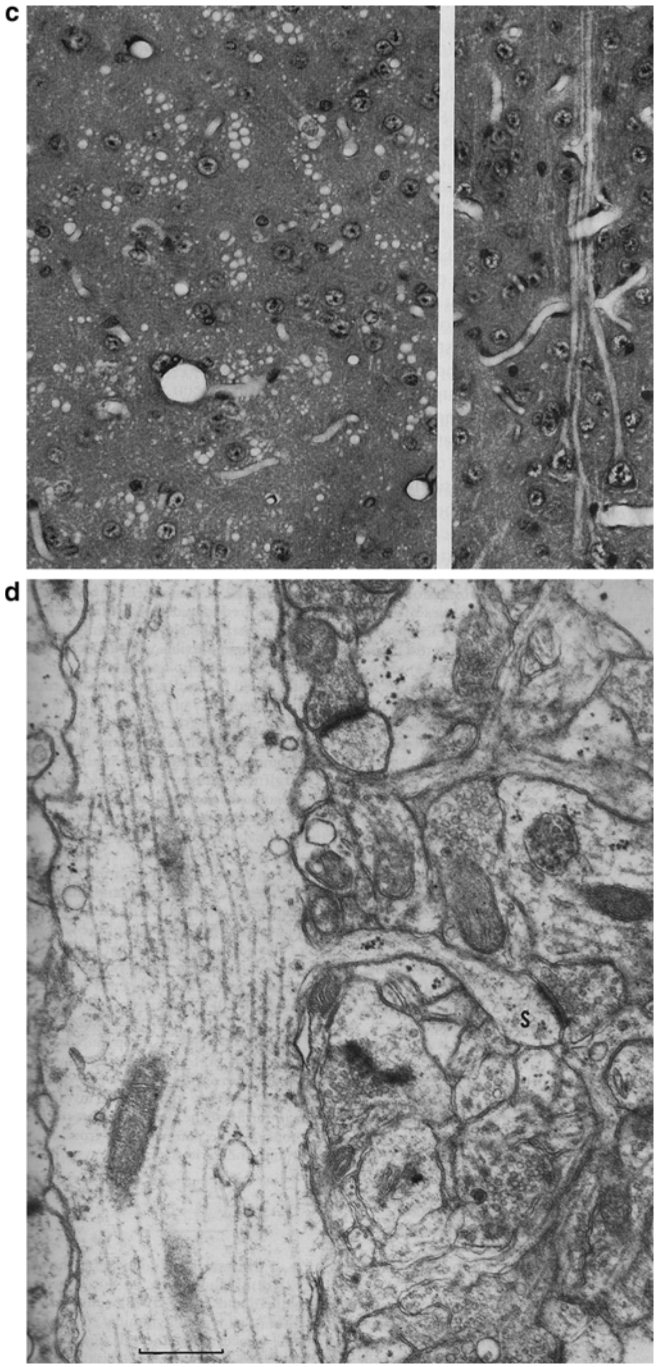


Fig. 14.1 (continued) The vertical filaments in the dendrite are axially oriented microtubules. The calibration line at the *bottom* of the micrograph is 0.5 µm ((**a**) and (**b**) from Eccles 1994, p. 128; (**c**) from Fleischhauer 1978, p. 106; (**d**) from Feldman 1984, p. 151. All with permission) (Color figure online)

Does Eccles' neuroquantology make sense? When did mind first appear in evolutionary history? Does it save 'mental indispensability' from the reflexologists? If so, does his theory have any evidential base in neurophysiology?

Eccles' position on the first question varied during his scientific life. In his 1986 paper to the Royal Society he implies that the mind somehow pre-exists the brain. In the nineteenth century Maudsley had rather the same idea when he wrote of a 'mentiferous aether' into which brains were able to tap.[69] Eccles reasoned similarly. He suggests that, when animal brains have reached an adequate level of complexity, they are able to pick up a pre-existing mental 'field', rather as radio receivers can detect radio waves. Later, in his final writings, he retreats a little from this idea of a pre-existing mentiferous aether, an idea which must remind classicists of the ancient idea of an all-pervading enlivening pneuma.[70] In his final publication he writes, instead, that '…the mind-world came to exist when the evolving cerebral cortex had microsites with synaptic vesicles poised in the presynaptic grids'.[71]

Eccles is ambiguous about the ontological status of his 'psychon fields'. In his 1986 paper and in other places he writes that they are 'analogous to quantum probability fields…'.[72] Elsewhere he seems to regard them simply as quantum fields. If psychon fields are immaterial, mental, '*res cogitans*', they provide, for Eccles, at long last, the 'stuff of consciousness' and save 'mental indispensabilty' and will's 'freedom'. If they are merely variants of the physicist's quantum mechanics, then Eccles' theory is materialistic and the immaterial mind still has no place in neuroscience.

The third issue is yet more problematic. For, since Eccles wrote his last papers in the 1980s and early 1990s, huge advances have been made in synaptology. It has become apparent that synapses are far more various than was apparent in Eccles' day. They are made not only between axon and dendrite and axon and perikaryon, but also between axon and axon, axon and glia and, according to a recent report, even between axon and myelin.[73] They are frequently grouped in intricate complexes. Their ultrastructure has been revealed as exceedingly complex, down to the molecular level. The processes of exocytosis, whereby neurotransmitters and neuromodulators are released into the synaptic gap, are still being researched but are already understood to be the outcome of teams of intricately constructed proteins. The biochemistry of the synaptic gap is also multifarious and we have not yet reached the post-synaptic density where further intricate biochemistry awaits. This is not the place to review all this molecular biology. It can be said, however, that it seems overwhelmingly likely that any sub-atomic quantum effects would be completely swamped in this manifold complexity. The presynaptic grid, moreover, upon which Eccles placed so

[69] Maudsley 1870. William James and Gustav Fechner also argued for something rather similar at the end of the nineteenth and beginning of the twentieth century. See Chap. 13 this volume.

[70] See Smith et al. 2012.

[71] Eccles 1994, p. 118.

[72] Eccles 1986.

[73] Stys 2011.

much store, can be found in organisms he would have regarded as untouched by the problems of free will – the Mauthner neurons of teleost fish, for instance, and in the sea hare, *Aplysia*.

Eccles concludes his final book, *How the self controls its brain*, in which he reviews his scientific odyssey, with the observation that his theory 'redeems what would otherwise be a mindless world, peopled by unconscious beings.'[74] He also refers to his book as chronicling 'the romance of his life'. It is thus melancholy to conclude that, after so great and sustained an effort to bridge the mind-brain crevasse, his theory remains only wishful thinking. Perhaps, though, it does show by way of an ultimate *reductio ad absurdum* that the division between the two sides of our world is unbridgeable within our current weltanschauung or world picture. As for Eccles, we can say of him that he never gave up and that he fully deserves Tennyson's encomium: 'Old age hath yet his honour and his toil/Death closes all: but something ere the end, some work of noble note may yet be done/… Tho' much is taken, much abides; and tho'/we are not now that strength which in old days/Moved earth and heaven: that which we are we are/… strong in will/to strive, to seek, to find, and not to yield.'[75]

14.1 Concluding Remarks

In this chapter we have traced the influence of Descartes' decisive separation of *res cogitans* from *res extensa* through the work of some of the greatest names in British neuroscience. We have seen that this dichotomy, originating in the heat of the seventeenth-century scientific revolution and, arguably, designed to save the ecclesiastical authorities from themselves, has proved unbridgeable. More than this, while knowledge of the *res extensa* side of the divide has increased exponentially, that of the *res cogitans* side remains, as Sherrington remarked, hardly developed at all. We are faced with an increasing take-over of that second side by the first: neuroimaging, molecular and evolutionary neurobiology are often, without thought, regarded as explaining 'mind'. Yet, on further analysis, the 'hard problem' emerges and remains as hard as ever.

We have seen in this chapter how efforts to account for 'freedom of the will' ultimately seemed to have run into the sand of neuroquantology. Perhaps, indeed, the will is not free and those who believe it to be labour under an illusion. Libet and his followers have made a strong case for this conclusion.[76] Palaeanthropologists might well back the Libetian neurochronology. They might argue that our innermost conviction that the will is 'free' is indeed an illusion. A necessary illusion. For without it and its derivative notions of praise and blame, how could human societies

[74] Eccles 1994, p. 182.

[75] Tennyson, *Ulysses*. In Tennyson 1911, p. 165.

[76] Libet 2002.

operate? Perhaps its deep embedment in our self-understanding merely reflects the fact that social life lies at the root of humanity's evolutionary success.

It is, however, far more difficult to explain away the 'raw feels', the 'qualia', we all live through in our day-to-day lives! Tell the sufferer that the pain of his aching tooth is illusory! Here is and remains the 'hard problem'. If Descartes' dichotomy and its consequences grew from the seventeenth-century scientific revolution with its materialist and mechanistic subtext, perhaps in the twenty-first century a new understanding of the physical world will show that the problem, as Foucault half hinted, was but the outcome of a particular world view or, to use his terminology, a particular 'epistemological space'.[77] Perhaps Eccles and his fellow neuroquantologists,[78] although deeply unfashionable today and possessing little or no traction in contemporary neuroscience, are nevertheless pointing in the right direction. Was it not Nietzsche who, in another context, noted that when the entire horizon is swept away, or perhaps especially when it is, an unconscionable time often elapses before the full import of the deed is understood?[79]

References

Anderson P, Lundberg A (1997) John C Eccles (1903–1997). Trends Neurosci 20:324–325

Barrow JD, Tipler FJ (1988) The anthropic cosmological principle. Oxford University Press, Oxford

Beck F (1996) Can quantum processes control synaptic emission? Int J Neural Syst 7:343–353. Workshop on the control of random events in biological systems

Beck F (2000) Modern physics opens the pathway to understand mind-brain interaction. In: Eccles H, Biersack HJ (eds) Sir John Eccles: in memoriam – a tireless warrior for dualism. Ecomed, Landsberg

Beck F, Eccles JC (1992) Quantum aspects of consciousness and the role of consciousness. Proc Natl Acad Sci U S A 89:11357–11361

Borck C (1999) John C Eccles (1903–1997); neurophysiologist and neurophilosopher. J Hist Neurosci 7:76–81

Bunge M (1980) The mind-body problem: a psychobiological approach. Pergamon, Oxford

Denny-Brown D (1939) Selected writings of Sir Charles Sherrington. Hamish-Hamilton, London

Desmond A (1997) Huxley: evolution's high priest. Joseph, London

Eccles JC (1951) Hypotheses relating to the mind-brain problem. Nature 168:53–57

Eccles JC (1953) The neurophysiological basis of mind. Clarendon Press, Oxford

Eccles JC (1977) My scientific odyssey. Annu Rev Physiol 39:1–18

Eccles JC (1986) Do mental events cause neural events analogously to the probability fields of quantum mechanics? Proc R Soc B 227:411–422

Eccles JC (1989) Evolution of the brain. Creation of the self. Routledge, London

Eccles JC (1990) A unitary hypothesis of mind-brain interaction in the cerebral cortex. Proc R Soc B 240:433–451

Eccles JC (1994) How the self controls its brain. Springer, Berlin

[77] Foucault 2002, p. xi.

[78] See Chap. 18.

[79] Nietzsche 1882, §125.

Eccles JC, Gibson WC (1979) Sherrington: his life and thought. Springer, Berlin
Feldman ML (1984) Morphology of the neocortical pyramidal neuron. In: Peters A, Jones EG (eds) Cerebral cortex, vol 1. Plenum, New York, pp 141–145
Fillenz M (2012) Memories of John Eccles. J Hist Neurosci 21:214–226
Fleischhauer K (1978) Cerebral architectonics. The last fifty years and some problems today. In: Brazier MAB, Petsche H (eds) Architectonics of the cerebral cortex. Raven Press, New York
Foster M (1901/1970) Lectures on the history of physiology during the sixteenth, seventeenth and eighteenth centuries. Dover, New York
Foucault M (2002) The order of things. Routledge, London
Fuller J (2007) The poetry of Sir Charles Sherrington. Brain 130:1981–1983
Galilei G (1638) Discorsi …. Intorno a Due Nuove Scienze. Leida. Trans: Crew H, de Salvio A (1953) Dialogues concerning two new sciences. Dover Books, New York
Gaukroger S (1995) Descartes: an intellectual biography. Clarendon Press, Oxford
Gotfredsen E (1950) A dissertation on the anatomy of the brain by Nicolaus Steno. NYT Nordisk Forlag Arnold Busck, Copenhagen
Honderich T (1988) A theory of determinism: the mind, science and life-hopes. Clarendon Press, Oxford
Huxley TH (1874) On the hypothesis that animals are automata and its history. In: Huxley TH (1898) Collected essays, vol 1. Macmillan, London
Jack JJB, Redman SJ, Wong K (1981) The components of synaptic potentials evoked in cat spinal motor neurons by impulses in single group 1a afferents. J Physiol 321:65–96
Kermode F (1971) Modern essays. Fontana Press, Glasgow
Libet B (2002) The timing of mental events: Libet's experimental findings and their implications. Conscious Cogn 11:291–299
Margenau H (1984) The miracle of existence. Ox Bow Press, Woodbridge
Maudsley H (1870) Body and mind: an enquiry into their connection and mutual influence. Macmillan, London
Nietzsche F (1882) The gay science (trans and ed: Kaufmann W. 1974). Vantage, New York
Sherrington CS (1906; 2nd edn. 1947) Integrative action of the nervous system. Cambridge University Press, Cambridge
Sherrington CS (1925, 2nd edn, 1940) The assaying of Brabantius and other verse. Oxford University Press, London
Sherrington CS (1933) The brain and its mechanism. Cambridge University Press, Cambridge
Sherrington CS (1946) The endeavour of Jean Fernel. Cambridge University Press, Cambridge
Sherrington CS (1951) Man on the his nature. Penguin Books, Harmondsworth
Shetty P (2012) The numbers game. Nature 484:S14
Smith CUM (1982) Evolution and the problem of mind: 1. Herbert Spencer. J Hist Biol 15:55–82
Smith CUM (2009) The 'hard problem' and the quantum physicists. Part 2: modern times. Brain Cognit 71:54–63
Smith CUM, Frixione E, Finger S, Clower W (2012) The animal spirit doctrine and the origins of neurophysiology. Oxford University Press, New York
Stensen N (1669) Discours sur l'anatomie du cerveau. Nicolai Stenonis Opera Philosophica, Copenhagen, p 18
Stys PK (2011) The axo-myelinic synapse. Trends Neurosci 34:393–400
Tennyson A (1911) Poems of Tennyson. Oxford University Press, London

Chapter 15
Is There a Link Between Quantum Mechanics and Consciousness?

Barry K. Ward

"I think I can safely say that nobody understands quantum mechanics".

Richard Feynman in *The Character of Physical Law*.
Modern Library, 1994

15.1 Introduction

This essay examines the link (if any) between consciousness and quantum mechanics. A short history of quantum mechanics and a description of the 'double slit' experiment is presented in Sects. 15.1.1 and 15.1.2, respectively to enable the reader to grasp an understanding of the problems associated with quantum measurement. The 'Measurement Problem' that arises from the 'Copenhagen Interpretation' of quantum mechanics is presented in Sect. 15.2 and the need for a conscious observer to collapse a wave function is discussed in detail. The paradoxes of 'Schröedinger's Cat' and 'Wigner's Friend' are also examined in Sect. 15.2.1. Both of these paradoxes suggest a link between human consciousness and the quantum realm and are still a source of active debate among physicists, philosophers and neuroscientists.

Alternative interpretations of quantum mechanics such as the 'Heisenberg-Dirac Propensity Interpretation', Everett's 'Relative State' or 'Parallel Worlds' interpretation and Bohm's 'Hidden Variables' interpretation are discussed in Sect. 15.3. The role of consciousness in the 'Parallel Worlds' interpretation is discussed in greater detail along with four variations due to Squires, Deutsch, Lockwood and Albert and Loewer. A more detailed review of quantum theories of mind due to Stapp, Hodgson, Penrose and Eccles follows and the quantum field theory of mind due to Ricciardi, Umezawa, Freeman and Vitiello is discussed in Sect. 15.4.

B.K. Ward (✉)
Macquarie Centre for Cognitive Science, Macquarie University, NSW, Australia
e-mail: barryw33@dodo.com.au

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_15,

The Penrose and Hameroff theory of mind is examined in Sect. 15.5 together with the proposal that the brain is a quantum computer and the suggestion by Penrose that Libet's backwards referral result (where cortical activity in response to a stimulus must proceed for 500 ms to elicit a conscious response) can only be explained by retro causation in the brain.

Decoherence theory is the study of how the interference effects due to the superposition of quantum mechanical states are suppressed. The same process that is responsible for the suppression of interference effects in the quantum world is also responsible for the suppression of interference effects in macroscopic objects in the physical world but the extremely short decoherence time means that such effects are not observable for a long enough period to be detected. For example a particle larger than a 1 g mass has a decoherence time of only 10^{-23} s. The relationship between decoherence and wave function collapse and the observer problem is discussed in detail in Sect. 15.6. The suppression of interference terms and the effect of decoherence time on neural and sub neural events is also examined and the meaning of decoherence for collapse theories of mind is also investigated.

15.1.1 A Brief History of Quantum Mechanics

In 1900 and 1901, Planck[1] came up with a theoretical derivation of the Stephan-Boltzmann equation,[2,3] which describes the radiation emitted by a 'black body' (a black body absorbs all light that is incident upon it and also acts as a perfect emitter when the surface temperature is raised). Planck suggested that the total energy emitted was made up of elements of energy called quanta. Planck hypothesized that the energies of the atoms of a black body radiator are similar to the energies in a harmonic oscillator such as a pendulum, which are restricted to certain values. Each energy level is an integral multiple of a basic unit and is directly proportional to the frequency of the oscillator. The constant of proportionality between the energy and the frequency is called Plank's constant 'h' and has the value of 10^{-34} joule-seconds.

In 1905 Einstein[4,5] examined the photoelectric effect where electrons are ejected from the surface of certain metals when light of a particular threshold frequency is incident on the surface. Maxwell's wave theory of light did not explain the photoelectric effect and Einstein proposed a quantum theory of light to solve this. Einstein realized that Planck's theory made implicit use of a light quantum hypothesis and Einstein suggested that light of a certain frequency also has a certain quantized energy. Only light with sufficient energy would be able to eject

---

[1] Planck 1901.

[2] Stefan 1879.

[3] Boltzmann 1884.

[4] Einstein 1905.

[5] Einstein 1906.

Fig. 15.1 Continuous spectra of solids (*above*) and line spectra (*below*) for hydrogen

electrons from a metal surface. More intense light (more quanta) only results in more electrons and not more electrons at higher energy. Einstein suggested that light at long wavelengths (low energy) would not be able to cause the emission of any electrons, as the light would not have enough energy to break electrons away from the surface. Different metals were found to have different energy thresholds at which emission would occur. Einstein proposed that light is emitted, transmitted, and absorbed as particles he called 'photons'. The photon energy was dependent on the frequency of the light.

In 1913, Bohr[6,7,8] used quantum theory to explain both atomic structure and atomic spectra. Bohr derived the relation between the electrons' energy levels and the frequencies of light given off and absorbed and explained the structure of narrow light and dark bands found in atomic spectral lines, (see Fig. 15.1 above), however, Bohr's theory did not explain why some energy changes are continuous and some are discontinuous and there was no explanation of how an electron knows when to emit radiation.

During the 1920s, the final mathematical formulation of the new quantum theory was developed when Louis de Broglie[9] proposed that light waves sometimes exhibit a particle nature as in the photoelectric effect and atomic spectra, and at other times light waves may exhibit a wavelike nature as in the double slit interference experiments (see 15.1.2). This "matter-wave" hypothesis was later confirmed in 1927 by Davisson and Germer,[10] who observed wave–like effects in a beam of electrons.

Two different formulations of quantum mechanics were proposed independently by Erwin Schrödinger and Werner Heisenberg following de Broglie's suggestion. The first of these was "wave mechanics" due to Erwin Schrödinger.[11] This formulation uses a mathematical function called a 'wave function', which is related to the probability of finding a particle at a given point in space. Quantum systems can exist in this undetermined state until observed. The act of observation (or measurement) collapses the wave function into one particular stable state.

Bohr believed that the wave function represents our knowledge of the physical phenomenon we are studying, not the phenomenon itself. The wave function contains potentialities which are actualized or realized when an observation is made.

---

[6] Bohr 1913a.

[7] Bohr 1913b.

[8] Bohr 1913c.

[9] De Broglie 1924.

[10] Davidson and Germer 1927.

[11] Schröedinger 1926.

The observation causes the wave function to "collapse" into an actual manifestation and not a potentiality. This later became known as the "Copenhagen interpretation" of quantum mechanics. Problems and paradoxes associated with this and other interpretations will be discussed in greater detail.

An alternative mathematical formalism called "matrix mechanics" was developed by Werner Heisenberg.[12,13] This theory does not use a wave function but was shown to be mathematically equivalent to Schrödinger's theory. Heisenberg wrote his first paper on quantum mechanics in 1925 and in 1927 stated his "uncertainty principle". The uncertainty principle states that the process of measuring the position of a particle disturbs the particle's momentum and the process of measuring the momentum of a particle disturbs the particle's position so that the knowledge of a particle's position or momentum are mutually exclusive events.

The uncertainty principle places an absolute limit on the accuracy of measurement and as a result, the prior assumption that any physical system could be measured exactly and used to predict future states was abandoned. By combining Planck's constant, the constant of gravity, and the speed of light, it is possible to create a quantum of length (approx 10^{-35} m) and a quantum of time (approx. 10^{-43} s), called Planck length and Planck time, respectively.

15.1.2 The Double Slit Experiment

The double slit experiment was first carried out by Thomas Young[14] in 1804 and demonstrated the wave nature of light, which was previously believed to have only a particle nature. Young actually used the edge of a thin card to show interference effects, which is equivalent to the double slit arrangement shown below. (Note: You can carry out the same experiment by placing your forefingers together and observing a light source between the gap of your fist and second knuckles, you will see vertical bands due to interference.)

If we consider a wall with two narrow slits and a source of small indestructible balls that are fired at two slits as shown in Fig. 15.2, below. The wall behind the slits is impacted by any of the balls that pass through the slits. The distribution of balls on the screen indicates that any ball that was initially behind a slit passed through that slit.

If we now consider a source of monochromatic light waves with the same slit setup and a fluorescent screen we find that each slit becomes a new source of light waves and when these two light waves combine on the screen they interfere with each other and result in an interference pattern composed of dark and light bands as shown in Fig. 15.3, below. The dark bands represent destructive interference and the light bands represent constructive interference.

[12] Heisenberg 1925.

[13] Heisenberg 1926.

[14] Young 1804.

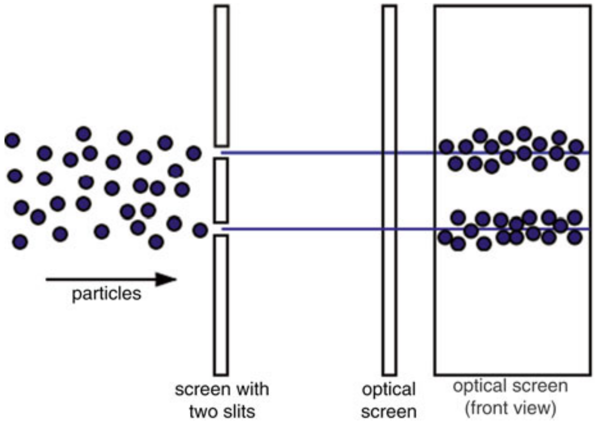


Fig. 15.2 Pattern produced for solid balls

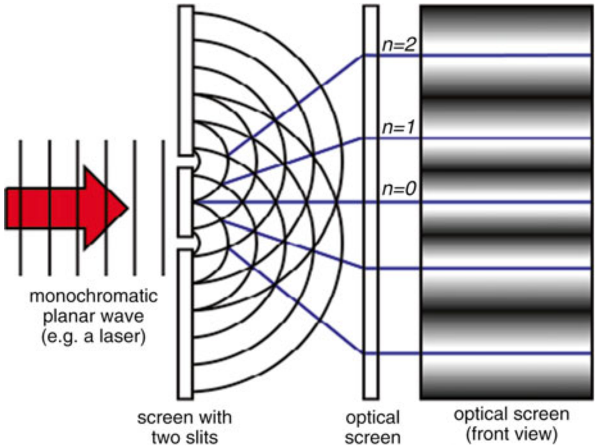


Fig. 15.3 Interference pattern for light waves

If the same experiment is carried out with electrons as a source we would expect the result shown in Fig. 15.2 due to the particle nature of electrons but because electrons also have a wave nature we end up with the interference pattern shown in Fig. 15.4, below.

If we now set up an experiment where we shine light on the electrons to determine which slit they are coming through then the resultant pattern is the same as shown in shown in Fig. 15.2. The act of observation and the knowledge that the electron has passed through one slit or the other destroys the interference pattern. If we close one of the slits then we get half of the solid ball pattern in Fig. 15.2.

If we now carry out the same experiment with a photon source that is limited to only one photon at a time over a period of days or months, then we still get the interference pattern shown in Fig. 15.4. The only explanation is that the photon has

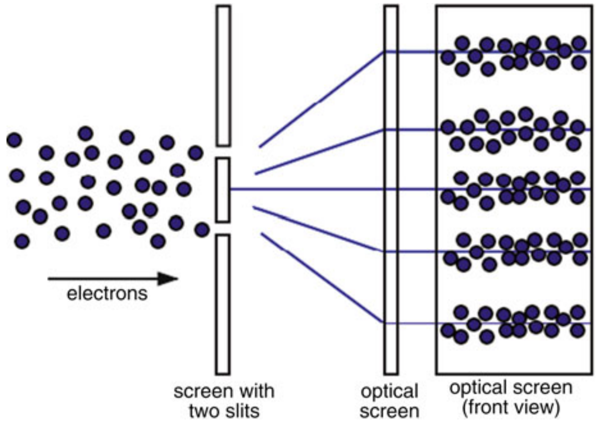


Fig. 15.4 Interference pattern for an electron source

the ability to interfere with itself. This behavior has also been observed with single electrons. The apparent 'wave–matter' duality of photons and electrons can only be explained with the aid of quantum mechanics.

15.2 The Measurement Problem

The quantum measurement problem came about as a result of the Copenhagen interpretation of quantum mechanics due to Neils Bohr.[15,16] The consensus at this time was that the time dependant Schrödinger equation predicts quantum states with many alternatives contained in a superposition of states. However, these alternative states are never observed (or actualized). Bohr's interpretation creates a division between the quantum world and the classical world and any measurement can only be performed with a classical apparatus. Bohr also proposed that this dividing line was not fixed and could in principle extend even to the human brain. The only requirement is a suitable measuring apparatus.

Quantum mechanics is a mathematical framework that describes the behavior of light and matter on the molecular, atomic and sub-atomic levels. Quantum theory has had many successful predictions and is considered to be the basis for all of physics but one aspect of the theory has remained unsolved for over 60 years. This problem, known as *'the measurement problem'*, is that the conditions for the actualization of potentialities (a superposition of quantum states) are not explicit in the formalism of quantum mechanics. That is, there no well defined physical or non-physical process responsible for the reduction of a superposition of quantum states (or potentialities) to a particular quantum state (or actuality). For example,

[15] Bohr 1928.

[16] Bohr 1935.

if a wave function represents the statistical probability of a particle's being observed, then a 'measurement' is said to 'localize' the particle otherwise the position of the particle is indeterminate.

Two very different and somewhat *ad-hoc* transformations occur in quantum mechanics, the first is a deterministic transformation of the wave function in accordance with Schrödinger's equation.[17] The second is a probabilistic transformation where the wavefunction undergoes a change from a pure state to a mixed state, which can only take place during a measurement (a 'pure' state is a linear superposition of all possible states and a 'mixed' state is one of all of those possible states). The problem is that the second type transformation is incompatible with the first type of transformation and Schrödinger's equation. Quantum physics (ie; Schrödinger's equation) applies to a quantum system up to the moment that a measurement is performed and classical physics applies to the measurement result. Quantum theory cannot explain how classical, physical phenomena emerge from quantum phenomena.

Another major problem with quantum theory is the notion of what constitutes a measurement. Is the dissociation of one molecule sufficient? Can a single photon perform a measurement or do we need a larger, macroscopic physical system. There is nothing in the formalism of quantum mechanics that defines what it is that constitutes a measurement. There is no clear demarcation between the macroscopic classical world of the measured state and the microscopic quantum world of the unmeasured state. The measuring apparatus is also subject to the laws of quantum mechanics regardless of the size or complexity of the apparatus. Maxwell[18] suggests that one solution to the above problem is to consider that the second type transformation can only occur when an observer becomes conscious of the result of the measurement.

This hypothesis proposes that a wave function is collapsed or reduced by some non-physical interaction between the consciousness of a human observer and the quantum system being examined as any physical interaction can just be considered as being part of the apparatus used to carry out the measurement. Maxwell[18] was not happy with this and considered it "*bizarre in the extreme that a purely physical process should occur only in those systems that interact with conscious observers*". Maxwell also suggested that the notion of a conscious observer could include any self-aware primate.

What is the dividing line between a measuring device and the quantum system being measured? Mathematically, a quantum system is a complex wave function (or pattern) of superposed wavefunctions. The components of the superposed state produce a complex interference function (or pattern) that describes the quantum state of the system. Any measurement performed on the system causes the interference effects to cease and leaves the system in a definitive quantum state (or 'measured state'). This process is irreversible.

---

[17] Schröedinger 1935.

[18] Maxwell 1974.

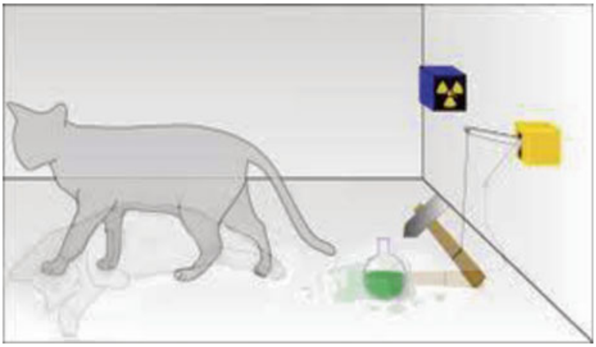


Fig. 15.5 Schrödinger's cat

Von Neumann[19] showed that during a measurement interaction, the combined system of object plus apparatus goes into a superposition of states where each state consists of the '*eigenstate*' of an observable together with a distinct state of the measurement apparatus (e.g.; a physical pointer on a dial). When an observer looks at the pointer, images from back of the retina form electrical impulses that travel to the visual cortex in the brain and become correlated with states of the combined system. Von Neumann suggests that this process can be extrapolated to include a correlation with the consciousness of the observer. The observer's consciousness will then go into a superposition of states where each state corresponds to a particular disposition of the measuring apparatus (e.g.; a pointer on a dial). Although Von Neumann's theory does not indicate exactly where and when the reduction of the wave function actually occurs he does suggest that this should occur no later than the registration of the measurement in the consciousness of the observer.

15.2.1 The Paradoxes of Schrödinger's Cat and Wigner's Friend

The paradox of Schrödinger's cat was published by Schrödinger[17] in 1935 to show that the description of a wave function was incomplete. Schrödinger suggested the following scenario (see Fig. 15.5): a cat is sealed in a box and a radioactive source is used to trigger a hammer that breaks a bottle containing cyanide killing the cat. The cat is said to be in a quasi alive-dead state until an observer opens the box and reduces the wave function to either a cat-dead or a cat-alive state (It should be noted that Schrödinger regarded this as a feature of description rather than an actual physical event). The paradox is that the cat is apparently in a linear superposition of cat-alive and cat-dead states until the box is opened by an observer and the wave function collapses to reveal one or the other observable states.

[19] Von Neumann 1955.

A second paradox was suggested by Wigner,[20,21,22,23] who used a ‘friend’ of the experimenter in place of the cat, and a light globe in place of the cyanide that is placed in a box. The light is turned on when a radioactive particle is detected and the ‘friend’ is instructed to report his observation to the experimenter. When the friend opens the box the larger wave function of the source, detector, light and friend will be reduced to a light on or light off state. If a second observer is introduced then he will collapse the wave function consisting of the source, detector, light, friend, and first observer. Wigner concluded that this process would lead to an infinite regress. Because of this paradox, Wigner concluded that human consciousness must be involved in the collapse of the wave function otherwise we end up with conscious observers in a multiplicity of states due to uncollapsed wave functions. If the first observer is conscious then he or she will collapse the wave function prior to the observation (or enquiry) of the second observer.

Although Schrödinger did not consider the cat paradox as a real physical situation, it was considered to be a serious problem by his colleagues and a long line of scientists and philosophers to this day. For example; Barrow and Tippler[24] proposed a number of ways to avoid the paradox such as Solipsism (the view that only oneself exists). They also suggest that any conscious being can collapse a wave function by observation and not just a trained observer and perhaps a ‘community’ of conscious beings can collectively collapse wavefunctions. They also proposed that an ‘ultimate observer’ may be responsible for the collapse of wavefunctions. Another proposal (that is shared with a number of alternative interpretations of quantum mechanics) is that the wave functions never collapse.

Hodgson[25] has three problems with the above proposals; Firstly, if it is the consciousness of one particular person that brings about the collapse of a wave function, then the wave function must be collapsed for all other conscious beings at the time of observation to avoid the observation of ‘different observables’ being observed. We then need a mechanism that explains how every conscious mind is connected. Secondly, there is the possibility of error in the observer due to say brain dysfunction or error due to the malfunction of the measuring apparatus and Thirdly Hodgson considers the case where two photographs of the observable are taken in succession. Does the observation of the second photograph immediately bring about the collapse of the first photograph?

Schröedinger himself suggested that wave function collapse occurred whenever a permanent record of the system was made and Heisenberg[26] proposed that

---

[20] Wigner 1961.

[21] Wigner 1962.

[22] Wigner 1967.

[23] Wigner 1977.

[24] Barrow and Tipler 1986.

[25] Hodgson 1991.

[26] Heisenberg 1958.

thermodynamic irreversibility was responsible. Davies[27] and Penrose[28,29] both propose that wave function collapse could only occur in the presence of a gravitational field.

15.3 Interpretations

Interpretations of quantum theory may be grouped into five main classifications: Bohr's 'Copenhagen' or 'Orthodox' interpretation, the Heisenberg-Dirac 'Propensity' interpretation, Everett's 'Many Worlds' interpretation, Bohm's 'Pilot Wave' interpretation and the 'Real Particle' interpretation. Each of these has a specific role for consciousness (the views of Bohr, Schrödinger, Heisenberg and Pauli on the 'hard problem' are summarized in Smith[30,31]).

15.3.1 Bohr's 'Copenhagen' or 'Orthodox' Interpretation

In 1926 and 1927 the Copenhagen interpretation of quantum mechanics was formulated by Bohr[15,16] and Heisenberg.[26] The most commonly accepted version is: "*Quantum mechanics is a tool for producing predictions rather than a theory for describing the world, whereas, classical terms have direct factual reference.....the classical level and the quantum level are entirely distinct and the transition from one to the other cannot be further analyzed*" (Feyerabend[32,33]).

Bohr and Heisenberg believed it was impossible to distinguishing between the objective and subjective at the quantum level and followers would write of the interaction between the observer and object causing large changes in the system under observation. It was commonplace to find expressions such as *'the observation disturbs the phenomenon' and 'the measurement creates the physical attributes of the object'*. Bohr[16] was to later change his view to a completely objective interpretation and suggested that "*it is not possible to conceive the quantum-mechanical state of an isolated microscopic system*", the system must include the measuring apparatus.

In Bohr's 'Copenhagen' interpretation, quantum theory does not describe a physical world that is independent of human observers. There is also some uncertainty in

27 Davies 2004.

28 Penrose 1986.

29 Penrose 1996.

30 Smith 2006.

31 Smith 2009.

32 Feyerabend 1968.

33 Feyerabend 1969.

the meaning of the term "wavefunction". The 'Copenhagen' interpretation is incomplete as it does not tell us how or when a measurement actually occurs. The theory requires a "*cut*" between the quantum system being measured and the classical system doing the measurement but does not say where the "cut" will occur. The sudden change to the wave function during measurement is just part of quantum theory. The reason for the change is not explained in the theory.

15.3.2 The Heisenberg-Dirac 'Propensity' Interpretation

Heisenberg's 'Propensity' interpretation of quantum mechanics, refers to things in nature as "*events*" and quantum theory specifies the tendencies or "*propensities*" for events to occur. In Heisenberg's ontology the wavelike properties of nature are embedded in expectation values of Heisenberg operators. The wavelike properties of nature are interpreted as objective tendencies for "*actual events*" to occur and the actual events correspond to the particle aspects of nature. Events are accompanied by change in the Heisenberg state of the universe due to wave function collapse.

According to Heisenberg[26] "*The observation itself changes the probability function discontinuously; it selects of all possible events the actual one that has taken place … the transition from the 'possible' to the 'actual' takes place during the act of observation. If we want to describe what happens in an atomic event, we have to realize that the word 'happens' can only apply to the observation, not to the state of affairs between two observations. It applies to the physical not the physical act of observation, and we may say that the transition from 'possible' to 'actual' takes place as soon as the interaction of the object with the measuring device, and thereby the rest of the world, has come into play; it is not connected with the act of registration of the result in the mind of the observer*".

Actual 'events' can be "*recorded*" *or embedded in an* enduring structure that enables re-verification of the event such as, the blackening of a photographic plate. This ontology allows the external world to carry on existing irrespective of human observation. Schrödinger's cat is either dead or alive and not in a quasi alive-dead state until an observer opens the lid.

Stapp[34] suggests that "… *the general shape of enduring physical objects, including all of their quasi-permanent marks and deformities are considered to be fixed by the ongoing flux of actual events. These fixed, quasi-stable features of objects, and similarly of biological organisms, provide a quasi-stable matrix of robust quantum properties around which the more transient quantum properties evolve. Thus, physical objects, and also biological organisms are considered to be 'really there'….even though the object or organism is interacting with its environment in a way that is violently disturbing huge numbers of non-robust degrees of freedom*".

[34] Stapp 1993.

15.3.3 Bohm's 'Pilot Wave' *or* 'Hidden Variables' *and* 'Real-Particle' *Interpretation*

The '*hidden variable*' interpretation considers quantum mechanics to be incomplete. Hidden variables or parameters describe how a range of discrete quantum states differ from each other. Bohm's '*Pilot Wave*' interpretation[35,36] allows classical deterministic laws to govern the evolution of the universe. Reality consists of quantum type wave functions and a classical world of particles and fields embedded in a '*quantum field*', (see Smith[31] for a non-mathematical account). The 'Real-Particle' interpretation of quantum mechanics is an interpretation postulated by David Bohm in which the existence of a non-local universal wavefunction allows distant particles to interact instantaneously. This interpretation posits that both wave and particle natures are real. The wave function of a particle evolves according to the Schröedinger equation. It assumes a single deterministic universe that evolves without the collapsing of wavefunctions when a measurement occurs. Bohm also established that the non-relativistic form of Schrödinger's equation is compatible with point particles provided that all such particles are linked simultaneously throughout the universe.

15.3.4 Everett's "Relative State" or "Parallel World's" ("Many Worlds"*) Interpretation*

Everett's "*Relative State*" interpretation[37] proposes that the actual physical world is radically different from that perceived by human consciousness. Everett claims that wavefunctions never reduce and that "*The wavefunction changes with time only and always in accordance with the Schrödinger equation*". External reality splits into many branches or 'worlds' where each world contains one of the many different results due to observation. In Everett's view making observations is equivalent to reducing wave functions. A human observer or human consciousness is not needed.

A major problem with other interpretations is that a quantum system, such as the Schrödinger cat example, suddenly jumps from a superposition of states to a particular state as a consequence of a measurement. In the "many worlds" interpretation there is no actual reduction to only one state as all states, which make up a superposed state, coexist. According to Everett the universe splits into a number of copies with each copy containing one of the superposed states. For example; Schrödinger's cat would split into two coexisting parallel worlds, one containing a dead cat and one containing a live cat. The observer would also split into two.

---

[35] Bohm 1952.

[36] Bohm 1990.

[37] Everett 1957.

DeWitt[38] expresses the process as follows: "*Every quantum transition taking place in every star, in every galaxy, in every remote corner of the universe is splitting our local world into myriad copies of itself.....here is schizophrenia with a vengeance*".

15.3.4.1 Problems with the Many Worlds Interpretation

Firstly, we now need to explain what process causes the split. Is the split due to the 'act' of observation using some physical force or field or is the split caused by the interaction of an observer's consciousness. If the latter then we are left with a variation of the original measurement problem where we need to explain how individual consciousness's can split both the universe and itself?

Secondly, Squires[39] asks what the probabilities in the wave function are now probabilities of. For example; in the many worlds theory the probabilities of observing a particular spin state are no longer the same probabilities as the Copenhagen interpretation as one 'Me' observes one spin state and another 'Me' observes another spin state.

Thirdly, Deutsch[40] suggests that "*different parallel universes may be linked by being part of a physical object*" and that "*physical reality is the set of all of the universes evolving together*" so that interference effects involve some sort of fusion of worlds. In a Young's two slit experiment each path is represented by different worlds before detection, which fuse to form one world when interference is detected.

Fourthly, how do we determine which state (or universe) the observer is in? Lockwood[41] suggests that we have the experience of being in all of these states but at any one time, we are only conscious of one of these states. Lockwood suggests that the current state is designated by consciousness and that only states that are shared eigenstates of a favored set of observables can be designated.

Fifthly, Hodgson[25] suggests that the many worlds hypothesis does not in fact answer the measurement problem. Hodgson suggests a Schrödinger's cat experiment where the probability of observing a dead cat is 0.01 rather than 0.5. The experiment is repeated 100 times which results in 2^{100} worlds where the observed results would be grouped around a probability of 50 % and not 1 %. This result appears to violate the statistical predictions of quantum mechanics.

We may overcome this problem by splitting into 99 worlds (each with a live cat) and one world with a dead cat, however, Hodgson suggests that to have the probability outcome determine the number of worlds would be difficult as probabilities include irrational numbers which would result in partial worlds. Lockwood[41] also expresses a similar suggestion and states "*what one would need is a continuous infinity of worlds, for each outcome, with a measure, in the mathematicians sense, that was proportional to the probability in question*".

---

[38] De Witt 1970.

[39] Squires 1990.

[40] Deutsch 1985.

[41] Lockwood 1996.

15.3.4.2 Variations to the Many Worlds Interpretation

Variation due to Squires

To avoid the problem of splitting mentioned above, Squires[39] proposes the existence of "selectors" that have "the power to select results for particular observations" and that the selection mechanism is human consciousness. Consciousness makes a random selection of what to observe and has no influence on the 'wavefunction'. Part of the wave function then becomes 'more real' and it is this part that is observed. This avoids the splitting into two 'me's' with each aware of a different result.

Everett[37] believes this modification to his theory "*does not have trajectories for the particles; indeed the external world does not even have particles; these are entirely a creation of conscious mind; like free will and redness, they are experiences*". There is also a problem when two observers each observe the same process but the detectors are separated by a large distance. If one observer makes a measurement and records say a particle in a plus spin state then the other observer will not see the particle according to Squire's variation. However, Quantum mechanics gives an equal probability of observing plus and minus spin states so how does the wavefunction inform the second observer that no particle is there if the first observer does not alter the wavefunction?

Squires[39] came to the conclusion that: "*It is with considerable hesitation that I suggest that the answer must lie in some sort of universal nature of consciousness*". Here, Squires refers to a universal mind through which individual minds interact but not at a conscious level. (This would need some sort of non-physical coupling between conscious individuals and a universal mind). This is a problem for any time period without consciousness as there could be no particle decay and the vacuum state of the universe (which fixes all physical parameters) would not exist and we must admit a degree of consciousness to every sub atomic particle and every rock and tree frog. Cochran (1971)[42] hypothesizes that the heat capacities of proteins may have a rudimentary degree of life.

Squires proposes that this problem can be eliminated if we consider that all past and present history of the universe is a subset of a much larger universal wavefunction that has been constantly evolving (equivalent to a quantum form of the '*Strong Anthropic Principle*').

Variation due to Deutsch

Deutsch[40] suggests that an infinite and constant number of parallel universes have always coexisted and their number remains constant. When a choice is made over a quantum event then the universes are partitioned into groups where one outcome occurs and groups where the outcome occurs. The universes increase in complexity in accordance with the second law of thermodynamics (This is a problem for quantum processes in biological systems which appear to develop against the rules of the second law).

[42] Lockwood 1996.

Variation due to Lockwood

Lockwood[41,43] suggests it is misleading to talk of physical worlds splitting as dividing or splitting in the relative state interpretation is equivalent to going into a macroscopic superposition. In the case of Schröedinger's cat, one could say that the universe as a whole is also in a superposition of cat-alive, cat- dead states, however, Lockwood[43,44] believes, "*Only in a Pickwickian sense could the rest of the universe be 'affected' by what befalls Schrödinger's cat*". Lockwood suggests that it is not the observer that splits the universe, it is the universe that splits the observer, as different *eigenstates* of a system become correlated with different brain states of the observer.

Lockwood's view implies that all human decisions are in fact indeterminate as all actions and their results become real alternatives. This would result in no real moral value for any action as (irrespective of the action) all alternative actions are realized (a similar problem occurs in the 'Orthodox Interpretation' where the reduction of 'potentialities' to 'actualities' involves random choice).

Variation due to Albert and Loewer

Albert and Loewer[45,46] propose a 'many minds' variation where any "*sentient physical system*" can take the part of an observer. This would involve an infinite set of minds and the "*array of choices embedded in the Schrödinger equation corresponds to the myriad of experiences undergone by these minds rather than to an infinitude of universes*".

15.4 Quantum Theories of Mind

The following is a brief review of quantum theories of mind due to Stapp, Penrose, Hogson, Eccles and Freeman and Vitiello.

15.4.1 Stapp's Theory

Stapp's proposal[34,47] is based on Heisenberg's picture of the physical world. Heisenberg suggested that atoms and electrons are not "*actual*" things such as a table or chair. The physical state of an atom or group of atoms or electrons is represented

[43] Lockwood 1989.

[44] Lockwood 1989.

[45] Albert and Loewer 1988.

[46] Albert and Loewer 1989.

[47] Stapp 2001.

by a set of "*objective tendencies*" or "*propensities*" for "*actual events*" to occur and these events can be measured or observed in the real physical world. These propensities or tendencies follow continuous deterministic mathematical processes, which obey the laws of classical physics. A second dynamic process brings about the occurrence of "*actual*" things in nature. This second process is termed a "*quantum jump*". Individual "quantum jumps" cannot be described by any physical theory but collectively they do obey statistical rules.

According to Heisenberg,[26] the deterministic part of quantum mechanics represents probabilities but the mathematical framework of quantum mechanics does not indicate what these probabilities are the probabilities of. Heisenberg suggested that these probabilities are "*objective tendencies*" for actual events to occur where the actual event is defined as "the actualization of one of the distinct metastable configurations of the observable degrees of freedom generated by the mechanical laws of motion, and the eradication of those remaining patterns of physical activity that might have been actualised, but were not". Stapp proposes that Heisenberg's picture couples quantum theory to an evolutionary description of physical reality and is not just a statistical set of rules about connections between human observers. I believe that Heisenberg's view only sidesteps the measurement problem as we are still left with the problem of how the potentialities are in fact 'actualized'.

Stapp[48] suggests that conscious events can be identified with physical brain events for the following reasons:

(a) Each nerve terminal in the brain exists in a mixture of quantum states. This is due to calcium ion precursors at synaptic junctions that require quantum theory to fully describe their behavior. Therefore, according to Stapp, the entire brain contains a cloudlike mixture of quantum states.
(b) Classical physics cannot explain consciousness without dualism which is not an issue if quantum mechanics is reintroduced into the problem.
(c) The decoherence time for ions (in aqueous solution) is much too short for quantum effects to play any significant role, however, Stapp suggests that the "quantum Zeno effect" can lengthen the decoherence time. The "quantum Zeno effect" occurs when the act of rapidly observing a quantum system forces that system to remain in an indeterminate state and prevents the system from collapsing into a particular, determined state. This effect is not diminished by the environment so that the decoherence time is extended. The simple observation of a quantum system suppresses certain of its transitions to other states. Stapp claims that the quantum Zeno effect is the main method by which the mind holds a superposition of the states of the brain in the process of attention. This is the principal method by which the consciousness can bring about change.
(d) Stapp proposes that each individual is equipped with three representations or schemas: a body schema used to execute bodily responses, an external world schema associated with the external world and a belief schema which is the current representation of a general historical schema. Projected self and world

[48] Stapp 2007.

schemas are selected by a conscious acts and are used to guide the organism. As these schemas may be manipulated by appropriate processing they are in a sense 'classical'. Stapp suggests that these schemas may be represented by physical structures in the brain and these structures are equivalent to observables in quantum mechanics.

Stapp[34] believes that the billions of synapses which are coupled together in a non-linear fashion should result in "a *huge number of metastable, reverberating patterns of pulses into which the brain might evolve*". Non-linear systems in the brain are sensitive to variations in input parameters (in this case, synaptic parameters). Synaptic processes are dependant on a small number of calcium ions resulting in a large number of metastable states into which a brain may evolve (see Smith[31] for more details). In the absence of quantum jumps, "*a brain will generally evolve quantum mechanically from one metastable configuration into a quantum superposition of many metastable configurations… that ascribes non-negligible quantum probabilities to several alternative possible metastable states of the 'self and world schema'"* Metastable patterns will become unstable due to the fatigue characteristics of synaptic junctions. The system will then be "*forced to search for a new metastable configuration, and will therefore continue to evolve, if unchecked by a quantum jump, into a superposition of states characterized by increasingly disparate self and world schema's*" (Stapp[34]).

Stapp maintains that a materialist theory will eventually account for consciousness but disagrees with Dennett's multiple drafts model (Dennett[49]), where the idea of a single stream of consciousness is an illusion. (Note that there is a great deal of evidence for fragmentation such as the Kolers-Grunau result, the Gray Walters precognitive carousel and Libet's subjective delay (Dennett[49])).

According to Stapp there are two factors that determine which alternative brain activities are actualized by an actual event. The first factor is local deterministic evolution governed by the Heisenberg (and Schrödinger) equation of motion. Historical influences such as learning and values may also influence tendencies associated with alternative courses of action.

The second factor selects one particular course of action from top-level patterns in the brain. This second factor according to quantum theory is chance. Stapp believes that "the basis for quantum choices cannot be conceptualized in terms of the ideas that it employs so that such choices appear to come from "*nowhere*" and must therefore be "*irrational*". This makes free will a problem for Stapp's hypothesis.

There is one further implication for the Heisenberg interpretation of quantum mechanics when applied to choices between distinct alternatives. Such choices are not due to local actions but are the result of global actions that transcend space and time (due to Bell's theorem[50]). Quantum theory predicts that "*although the flow of conscious events associated with a particular human brain has important personal*

[49] Dennett 1991.

[50] Bell 1987.

aspects, … the fundamental process that is expressing itself through these local events is intrinsically global in character" (Bell[50]). The term "pure chance" is used to describe this global process.

15.4.2 Hodgson's Theory

Hodgson[45] suggests that the mind and brain are "both manifestations of the same underlying reality" and that mind can be interpreted as an emergent function of the brain only if we assume there is an underlying quantum reality associated with both the physical brain and mental events of consciousness. Hodgson believes that the external world is the result of gross statistical properties of a cosmic code. We detect and interpret this code as sensory events and objects. Hodgson does not believe that mental events are the result of gross statistical properties of quantum events in the brain and proposes that mental events are related to quantum processes directly. Hodgson further suggests that associated with what appears to be an "*apparently unified and indivisible conscious experience … is a pattern of physical events which are substantially cotemporaneous and spatially extended.*" Perception of an object such as a red ball moving through the air involves the recognition of various features such as color, shape and movement and the comparison with previous beliefs about what the object is. These processes involve spatially and temporally extended regions of the brain but the subjective experience appears to imitate the physical character of the external world.

The fact that changes in our experience appears to be simultaneous (when presented to consciousness) can be explained by short-term memory. However, Hodgson suggests that short-term memory alone cannot explain the feeling of a specious present and our feeling of the passage of time. Contributions to experience from short-term memory may also involve neural events from spatially extended regions of the brain. From Hodgson's view we can never be "truly aware" of an external reality.

Some evidence of this is found in patients with short-term memory dysfunction (one such patient would write every 10 min, the statement that: *now for the first time I am truly aware*). It may well be that we can consciously experience external reality without associations with prior concepts from long-term memory and contributions from short-term memory. Such an experience would be without any categorizing or labeling.

It would appear that mental events somehow span space, enabling simultaneous experiencing of spatially separated physical events. "*Instantaneous correlations of spatially separated events are only found in the potentialities of quantum state*" thus Hodgson[45] believes it is "*plausible to associate mental events closely with the quantum physical states manifested by brain events*".

The integration of mental events to produce a collective *"wholeness"* underlies the hypothesis of both Hodgson and Stapp. However, this is undermined by the fact that consciousness may be due to a collection of conscious subsystems that are somehow integrated into a collective whole. A well known example is found in *split brain patients* with apparent dual centers of consciousness leading Nagel[51] to write "*If I am right, and there is no whole number of individual minds that these patients can be said to have, then the attribution of conscious, significant mental activity does not require the existence of a single mental subject.*"

Conscious subsystems are also found in the experiments of Libet, Feinstein and Pearl,[52,53] which show there is no conscious sensation (of stimulation to the skin) unless preceded by unconscious cortical activity for periods up to half a second. Libet suggests that the delay has the function of "*keeping ongoing sensory inputs from reaching conscious levels*" and provides "*an opportunity for modulating a perception*". Hodgson suggests that it may be possible for decisions to be made by conscious subsystems without our knowledge and that "*consciousness of such parts may at different times be (or be not) integrated into a single consciousness*". Examples are also found in patients with multiple personality disorders where each individual personality may be entirely controlled by a conscious subsystem.

15.4.3 Penrose's Theory

The core of Penrose's theory of consciousness (Penrose[29,54,55,56]) is that the shared *"global"* character of conscious thought is similar to a quantum state or quantum states. Examples are found in mathematics where conscious thought instantaneously grasps a complex whole. Penrose also believes that "*the action of conscious thinking is very much tied up with the resolving out of alternatives that were previously in linear superposition*" and these alternatives are similar in nature to superposed quantum states. The main difference between Stapp and Penrose is that Penrose considers this process to be non-computational and believes that "*appropriate physical action of the brain evokes awareness*" but "*this physical action cannot even be properly simulated computationally*" (Penrose[54]).

Penrose considers that quantum systems may evolve in two different ways. The first way is a deterministic "*unitary*" process (U) and the second way is a "*collapse*" or "*reduction*" process (R). The (R) process is a physical action that "*is non-local in a way that is consistent with the type of violation of Bell's inequality that has been observed in actual experiments*" and is non-computational. Penrose believes there is a similarity between consciousness processes and R-type processes and states, "*the phenomenon of consciousness are dependent upon some physical process that underlies the R-procedure of quantum mechanics*".

[51] Nagel 1976.

[52] Libet et al. 1979.

[53] Libet et al. 1992.

[54] Penrose 1989.

[55] Penrose 1994.

[56] Penrose 1997.

Penrose also believes that Libet's backwards referral mechanism is evidence for retrocausation in conscious thought and is subject to the laws of quantum mechanics (Libet found that cortical activity in response to a stimulus must continue for about 500 ms to elicit a conscious sensation. The timing of the sensation is then subjectively *referred* back to the initial stimulus), however, there is no scientific evidence for time-reversed processes in our brains that may or may not require quantum mechanics. It should be noted that Libet did not agree with Penrose's interpretation and believed that the backwards referral mechanism was an 'as if' situation rather than physical retrocausation.

Penrose together with Hameroff[57] suggest that microtubules may be plausible sites for quantum mechanical processes involved with consciousness. Microtubules lend structure and create pathways for chemical transport within nerve cells and computer models show that the insulating properties of microtubules may allow vibrational pulses to explore multiple pathways. The main problem with this hypothesis is how microtubules communicate with cells. The use of neuromodulators would require activity that is very large on a quantum scale. There is also the problem of a mechanism for quantum coherence in microtubule clusters and the requirement of a yet to be determined theory of quantum gravity essential to the whole theory. More recent findings in neurobiology by McKemmish, Reimers, McKenzie and Hush[58] suggest that "tubulins do not possess essential properties required for the Orch-Or proposal" and "recent progress in the understanding of the long-lived coherent motions in biological systems" indicate that coherent computations in microtubules is not possible.

15.4.4 Eccles's Early Quantum Theory of Mind

Eccles[59] proposed that quantum processes in brain dynamics and nerve terminals were the basis for the link between mind and brain. However, this approach introduces a bias to quantum statistics. Initially, Eccles[60] believed that quantum indeterminacy would take over at the microscopic scale in the brain. However, the magnitude of diffusion forces found in synaptic junctions were found to be much larger than any quantum effects (Beck[61]). Eccles[62] suggested that: "*mental events alter the probability of vesicular emission that is triggered by a presynaptic impulse*". Beck and Eccles proposed that exocytosis (the release of vesicular contents from a neuron) must be atomic in nature, i.e.; "an incoming nerve impulse excites some electronic configuration to a metastable level which is separated

[57] Hameroff 1994.

[58] McKemmish et al. 2009.

[59] Eccles 1990.

[60] Eccles 1986.

[61] Beck 1996.

[62] Eccles 1994.

energetically by a potential barrier from a state which leads unidirectionally to exocytosis" (Beck[61]). Quantum tunnelling through this potential barrier results in the generation of a superposed state of two wave functions representing the penetration or non-penetration through the barrier. Any 'act' of observation due to conscious choice collapses the wave function into one state or the other (exocytosis or non-exocytosis).

Eccles proposed that quantum processes in thousands of presynaptic membranes in cortical pyramidal cells resulted in unitary mental event or 'psychon'. Each psychon influenced the probability of exocytosis in all the synapses associated with a dendron. "Our mentality consists of a shifting mosaic of psychons each linked to a cortical dendron" (also see Smith[31]). Earlier, Eccles[60] suggested that mind is independent of brain due to a pre-existing "mental field" which is accessible to any brain sufficiently complex (He withdrew from this view at a later date).

There are three main problems with Eccles approach: firstly, there is no indication of how consciousness selects which state is to be realized. Secondly, recent research in neurobiology has yet to find any quantum like processes that may be responsible for the release of neurotransmitters into the synaptic cleft or to affect the initial trigger due to the influx of Ca^{2+} ions (see Smith[31] for a summary) and thirdly, for the "*hard problem*" we are left with a form of dualism.

15.4.5 Ricciardi, Umezawa, Freeman and Vitiello's Quantum Field Theory of Mind

Ricciardi and Umezawa[63] were among the earliest to suggested that quantum field theory may be applicable to brain states. Umewaza suggests that memory states are similar to the states of a many-particle system such as is found in the vacuum states of quantum fields. Umewaza proposes that the brain is a many-particle system and that neurons behave as particles. Coherent neuronal assemblies would then be analogous to the dynamically ordered states of a many-particle system and the encoded content of a neuronal assembly would be consciously accessible via an external stimulus. This allows the formation of memory states with a finite lifetime and conscious recall of content.

Vitiello[64,65] further considered the problem of dissipation from interaction with the environment and suggests that the doubling of collective modes (in the form of differently coded vacuum states of quantum fields) would enable the possibility of memory storage without overprinting. The affect of external stimuli on the stability of such states has been investigated by Stuart, Takahashi and Umezawa[66] and the

[63] Ricciardi and Umezawa 1967.

[64] Vitiello 1995.

[65] Vitiello 2001.

[66] Stuart et al. 1978, 1979.

affect of chaos and quantum noise have also been addressed by Pessa and Vitiello.[67] Vitiello, also proposes that a "time-reversed copy" of brain states may be possible so that "consciousness seems thus to emerge as a manifestation of the dissipative dynamics of the brain". In a later publication, Freeman and Vitiello[68] suggest that electric field amplitudes and neurotransmitter concentrations remain purely classical and do not require the application of quantum theory.

Vitello's theory of mind allows mental activity to be correlated with the dynamics of neuronal assemblies and avoids many of the restrictions associated with the standard version of quantum mechanics. However, there are two problems with this theory, firstly, if brain states are determined by quantum field theory then where is the neurobiological evidence for this? And secondly, the majority of presentations of this view do not distinguish between mental states and material states.

15.5 Penrose and the Brain as a Quantum Computer

One of Penrose's more speculative suggestions regarding human thought is that he believes human minds are non-algorithmic and therefore cannot be equaled by any form of artificial intelligence. Similar claims have been made by Godel[69,70] where Godel suggests "… the human mind (even within the realm of pure mathematics) infinitely surpasses any finite machine, or else there exist absolutely unsolvable diophantine problems". Penrose proposes that 'non-algorithmic' is the same as 'non-computable' in the sense that human thought cannot even be approximated by a formal operating system that is algorithmic. Searle[71] also believes that AI programs cannot 'think' in the same way that humans think irrespective of complexity but could possibly imitate consciousness, however, imitation doe not imply 'consciousness' (human consciousness).

Penrose believes that all mental processes are basically physical processes and that Godel's theorem that "no consistent algorithm can produce a proof of its own consistency" and "the totality of processes by which I can come to accept mathematical statements as true is either unknowable to me, or unsound" (Godel[70]). Penrose deduces from this that "Human mathematicians are not using a knowably sound algorithm in order to ascertain mathematical truth" (Penrose, 'Shadows of the Mind' 2.5). Penrose claims that behavior that imitates human consciousness will never be observed because of the reliance on algorithmic computation and its limitations that are not evident in our non-computational processing brains.

---

[67] Pessa and Vitiello 2003.

[68] Freeman and Vitiello 2006.

[69] Godel 1931.

[70] Godel 1951.

[71] Searle 1980.

Searle[71,72] argues that computation uses the manipulation of symbols, however, the symbols themselves are observer relevant and not part of reality ie; “Gravitational attraction, photosynthesis and electromagnetism are all subjects of the natural sciences because they describe features of reality, but the feature of being a bathtub or a five dollar bill exists only to observers and users”.

The question “Is consciousness a computer program” becomes “Can a computational interpretation be assigned to specific brain processes that characterize consciousness?” In other words nothing is intrinsically computational and “computation exists only for some agent or observer who imposes a computational interpretation on some phenomenon”. This implies that a computational model of consciousness cannot in itself be conscious (for example, the computational model of sitting in a bath of water does not leave us wet). I believe that Penrose makes the same mistake in his use of the Godel argument.

The argument that human mathematicians can come up with mathematical truths that cannot be proven through computation has been extensively debated for over 40 years without resolution (see Lewis,[73] Bowie[74] and Feferman[75]), however some experimental data suggests that human thought processes involving expert knowledge may be in part non-computational. Dreyfus[76] suggests, “It is not possible to capture expert knowledge in an algorithm, particularly where it draws upon general background knowledge outside the problem domain”. There has been limited success in building expert knowledge into rule-based machines but recent progress has seen artificial neural networks capable of learning and recognizing complex patterns. Such networks do not follow explicit rules but can be approximated by an algorithm; however, if Penrose is correct and human thought processes cannot even be approximated by an algorithm then artificial neural networks do not provide a counter argument.

Penrose’s arguments for non-computational human thought are at best vague and are an insufficient basis to propose that non-computational processes in microtubules (assisted by quantum gravity) are responsible for our inner subjective life. The experimental results from current neuroscience and the effects of decoherence must also be considered in any theory of mind. Koch and Hepp[77] suggest: “The critical questions we are here concerned with is whether any components of the nervous system – a 300 °K wet and warm tissue strongly coupled to its environment – display any macroscopic quantum behaviours, such as quantum entanglement, and whether such quantum computations have any useful functions to perform”.

[72] Searle 2007.

[73] Lewis 1979.

[74] Bowie 1982.

[75] Feferman 2006.

[76] Dreyfus 1972.

[77] Koch and Hepp 2006.

15.6 Decoherence

The main argument against large-scale macroscopic states and small-scale sub-cellular quantum states occurring in the brain is that the brain physiology is a wet, hot environment. Localized quantum states are prevented from linking or associating with other localized quantum states due to ‘decoherence’. Decoherence looks at the way a quantum system interacts with it’s immediate environment and in particular to the suppression of interference. A simple example of interference is a two-slit experiment (see 15.1.2) in which photons are fired at two narrow slits to a screen on the opposite side. Over a period of time an interference pattern emerges on the screen. If one photon at a time is used the interference pattern is still observed as a result of the photon interfering with itself. If one slit is covered or an act of measurement detects a photon at one of the slits then the interference pattern vanishes as only one component of the interference survives the measurement.

The time taken for the suppression of interference is termed the *decoherence time*. At the end of the decoherence time any coherence or phase relationships between components of the quantum system are destroyed. For example, the decoherence time for a 1 g mass at room temperature is less than 10^{-23} s and a dust grain interacting with background radiation in free space has only a few nanoseconds before any coherence is destroyed (Zurek[78]). The main issue for the formation of coherently linked quantum states in the brain is “whether the relevant degrees of freedom of the brain can be sufficiently isolated to retain their quantum coherence” (Tegmark[79,80]).

15.6.1 Decoherence Mechanisms in the Brain

The quantum-brain models examined previously rely on extended periods of coherence that approach classical neural processes. Stapp[34] suggests that some neural processes can be isolated from their environment whereas Zeh,[81,82] Zurek,[78] Tegmark,[79] Scott,[83] Hawking[84] and Hepp[85] argue that any quantum macrostates in the brain would be rapidly eliminated due to decoherence.

Decoherence times for typical sub-neural interactions were derived by Tegmark.[79,80] Tegmark found that decoherence times for ion-ion collisions was of

[78] Zurek 1991.
[79] Tegmark 2000a.
[80] Tegmark 2000b.
[81] Zeh 1970.
[82] Zeh 1999.
[83] Scott 1996.
[84] Hawking 1997.
[85] Hepp 1999.

the order of 10^{-20} s and for ion-water collisions was approximately 10^{-20} s and coulomb interactions with distant ions were found to have a decoherence time of approximately 10^{-19} s. Cognitive processes for speech, thought and visual processing have dynamic timescales of 1 s to 10^{-2} s. A single ion traversing a cell wall would have a decoherence time of approximately 10^{-14} s. This is obviously many orders short of the timescales associated with classical neural events and we are forced to accept the conclusion that any macroscopic neural or sub-neural event can be sufficiently explained using classical statistical mechanics.

15.6.2 Decoherence and 'Collapse' Approaches

It is useful to look at the role that decoherence plays in the collapse approaches to quantum mechanics due to Von Neumann and Penrose. Von Neumann[19] proposed that the collapse of a wave function is facilitated by an observer's consciousness. Collapse occurs whenever a permanent record is made in the visual cortex or the fluorescence on a screen or whenever consciousness is involved in an observation. Von Neumann assumes that there is an absence of interference between the components of the wave function. The presence of interference would affect the timing and the resulting classical outcome. For example, the collapse of the wave function in a two-slit experiment may occur anywhere from behind the slits to the screen. The reduction of any interference (decoherence) is thus essential to Von Neumann's collapse approach.

The Penrose and Hameroff[55] 'collapse' theory suggests that coherent superpositions of dimer states in microtubules can give rise to rise to excitations that travel along the dimmers at speeds greater than 1 m/s (Sataric, Tuszynski and Zakula[86]). Penrose and Hameroff believe that these long range coherent processes may act as a type of quantum computer in the brain and suggest that microtubules are the site of human consciousness ie; coherent superpositions in tubulin proteins give rise to a sub-conscious process neural event and the self-collapse of superposed states leads to a conscious neural event. In this 'Orch-OR' (Orhestrated Objective Reduction) model (see Smith[31]), the self-collapse is triggered by a (yet to be determined) quantum gravity mechanism (Penrose[29]). This is a type of "*pan-protopsychist*" solution to the 'hard problem'.

To prevent decoherence taking place there would be a requirement to maintain coherent superpositions of microtubule states for up to hundreds of milliseconds. Hagan, Hameroff and Tuszynski[87] claim that Tegmark did not look at superposed protein conformations which may extend the decoherence time to 10^{-5} to 10^{-4} s but the main problem with the 'Orch-Or' model is that any neural system that is isolated from the environment will eventually become 'conscious' if decoherence is prevented.

[86] Satari et al. 1993.

[87] Hagan et al. 2002.

15.6.3 Problems with Decoherence

A major problem with decoherence is to explain how a particular state is chosen in preference to another. We have two alternatives: either the system interacts with the immediate environment until probabilities associated with the system result in a collapse into one particular eigenstate. A measurement must be made to determine which state the system is in. As the system is already collapsed then the observer and the observation have no influence on the outcome. The system evolves without the help of a conscious observer. The other alternatively is to propose a decohered system that remains in a superposed state until a measurement is made and an outcome is observed. In this case it is the measurement itself that determines the outcome. Both situations predict that the system will be in one or another eigenstate.

Because of decoherence it is difficult to see how clusters of neurons, individual neurons or microtubules can exist in an extended, coherent linear superposition of quantum states at typical body temperatures. Any system larger than a molecule can be adequately described with classical probability calculus. It is the interaction between objects and their environment that brings about wave function collapse. The consciousness of an observer is unnecessary as the interaction with the environment rapidly destroys any coherent phase relationship between any macroscopically distinct states. The theory of decoherence can also be derived from within the formalism of quantum theory.

15.7 Conclusions

In my opinion, the role of the observer in quantum mechanics is still a matter of dispute. Any modification to the mathematical formalism is unlikely to improve the situation. The Copenhagen interpretation must accept the external world as physically "real", whereas, the "many worlds interpretation" (favored by Squires, Deutsch and Lockwood) provides a solution to the measurement problem but with a large amount of metaphysical baggage. The only "reasonable" variation of the "many worlds" theory that is closest to Everett's intentions is the "many minds interpretation" but this also comes with unresolved philosophical and scientific issues. The Heisenberg-Dirac interpretation, which is favored by Stapp, appears to sidestep the observer problem.

Large, high temperature items such as Wigner's friend, Schrödinger's cat, neurons and microtubules are unlikely to exist in a linear superposition of quantum states. Macroscopic systems are just not found in linear superpositions of coherent states and therefore may be adequately described by well-defined classical states. The paradox of Schrödinger's cat or Wigner's friend may be explained with the use of classical probability calculus and if this is the case then no observer is required to collapse the wave function.

I find Penrose's arguments unconvincing as cognitive studies show that formal reasoning in humans usually involves the use of heuristic shortcuts even amongst experts. There is also no evidence that physics is non-computable or that some yet to be determined quantum process is essential to cognition. Penrose and hameroff's view that awareness is the result of quantum computation in microtubules is difficult to accept and as Chalmers[88] in Psyche (1995) suggests; "..why should quantum processes in microtubules give rise to consciousness, any more than computational processes should?" Penrose also makes no mention of subconscious processing and argues that introspection must be conscious. However, the vast majority of mental processes are in fact subconscious (such as habituated stimuli, automatic skills and visual cognition).

The requirement due to decoherence that objects larger than a molecule cannot exist in a state of linear superposition is also a problem for the Penrose-Hameroff theory. The decoherence time to go from a superposition of states to a classically described state is orders of magnitude shorter than typical neuronal or sub-neuronal interaction times (this is also a problem for any theory of mind that requires quantum coherence such as the theories proposed by Stapp and Hodgson).

The concept of quantum type processes being responsible for higher brain function will remain a concept until validated by replicable experiments. It is more probable that a theory of brain function based on classical physics will adequately explain the integrative and holistic nature of conscious thought mentioned in the theories of Stapp, Hodgson and Penrose. A workable model may perhaps be found in future chaos or connectionist theories of mind. Regardless of choice, an understanding of consciousness will most likely require a similar paradigm shift found in the disciplines of philosophy, physics and neuroscience when confronted by Newton's gravity, Einstein's relativity and De-Broglie's 'matter-wave' hypothesis.

References

Albert DZ, Loewer B (1988) Interpreting the many worlds interpretation. Synthese 77:195–213

Albert DZ, Loewer B (1989) Two no-collapse interpretations of quantum mechanics. Noûs 12:121–138

Barrow JD, Tipler FJ (1986) The anthropic cosmological principle. Oxford University Press, New York

Beck F (1996) Can quantum processes control synaptic emission? Int J Neural Syst 7:343–353

Bell JS (1987) Speakable and unspeakable in quantum mechanics. Cambridge University Press, Cambridge

Bohm D (1952) A suggested interpretation of the quantum theory in terms of "hidden" variables. 'I' and 'II'. Phys Rev 85:166–179; 180–193

Bohm D (1990) A new theory of the relationship of mind and matter. Philos Psychol 3:271–286

Bohr N (1913a) On the constitution of atoms and molecules, part I. Philos Mag 26:1–24

Bohr N (1913b) On the constitution of atoms and molecules, part II. Systems containing only a single nucleus. Philos Mag 26:476–502

[88] Chalmers 1995.

Bohr N (1913c) On the constitution of atoms and molecules, part III. Systems containing several nuclei. Philos Mag 26:857–875

Bohr N (1928) The Copenhagen interpretation. Nature 121:580

Bohr N (1935) Quantum mechanics and physical reality. Nature 136:1025–1026

Boltzmann L (1884) Über die Eigenschaften Monocyklischer und andere damit verwandter Systeme. Crelles J 98: in WA III, paper 73. 68–94, (WAIII is Boltzmann L (1909) Wissenschaftliche Abhandlungen, vol I, II, and III, Hasenöhrl F (ed), Leipzig: Barth; reissued New York: Chelsea, 1969)

Bowie GL (1982) Lucas' number is finally up. J Philos Logic 11:279–285

Chalmers DJ (1995) Facing up to the problem of consciousness. J Conscious Stud 2:200–219

Cochran A (1971) Relationships between quantum physics and biology. Found Phys 1:235–250

Davidson CJ, Germer LH (1927) Diffraction of electrons by a crystal of nickel. Phys Rev 30:705–715

Davies P (2004) Does quantum mechanics play a non-trivial role in life? BioSystems 78:69–79

de Broglie L (1924) A tentative theory of light quanta. Philos Mag 47:446

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

Deutsch D (1985) Quantum theory, the Church-Turing principle and the universal quantum computer. Proc R Soc (Lond) A 400:97–117

DeWitt B (1970) Quantum mechanics and reality. Phys Today 23:30–35

Dreyfus HL (1972) What computers can't do. MIT Press, Cambridge, MA

Eccles JC (1986) Do mental events cause neural events analogously to the probability fields of quantum mechanics? Proc R Soc (Lond) B 227:411–422

Eccles CJ (1990) A unitary hypothesis of mind-brain interaction in the cerebral cortex. Proc R Soc (Lond) B 240:433–451

Eccles JC (1994) How the self controls its brain. Springer, Berlin/Heidelberg/New York

Einstein A (1905) On a heuristic viewpoint concerning the production and transformation of light. Ann Phys 17:132–148

Einstein A (1906) On the theory of light production and light absorption. Annalen der Physik (ser 4) 20:199–206

Everett H (1957) Relative states' formulation of quantum mechanics. Rev Mod Phys 29:454–462

Feferman S (2006) Are there absolutely unsolvable problems? Gödel's dichotomy. Philos Math 14:134–152

Feyerabend P (1968) On a recent critique of complementarity: part I. Philos Sci 35:309–331

Feyerabend P (1969) On a recent critique of complementarity: part II. Philos Sci 36:82–105

Freeman WJ, Vitiello G (2006) Nonlinear brain dynamics as macroscopic manifestation of underlying many-body field dynamics. Phys Life Rev 3:93–118

Godel K (1931) Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. Monatshefte für Mathematik und Physik 38:173–198

Godel K (1951) Some basic theorems on the foundations of mathematics and their implications. In: Feferman S (ed), 1995. Collected works, Kurt Gödel, vol III. Oxford University Press, Oxford, pp 304–323

Hagan S, Hameroff SR, Tuszynski JA (2002) Quantum computation in brain microtubules: decoherence and biological feasibility. Phys Rev E 65(061901):1–11

Hameroff SR (1994) Quantum coherence in microtubules: a neural basis for emergent consciousness. J Conscious Stud 1:91–118

Hawking S (1997) Objections of an Unashamed Reductionist. In: Penrose R (Author), Hawking S, Cartwright N, Shimony A (Contributors), Longair M (ed) The large, the small and the human mind. Cambridge University Press, Cambridge

Heisenberg W (1925) Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. Zeitschrift für Physik 33:879–893. (English translation in: van der Waerden BL (ed) (1968) Sources of quantum mechanics (English title: "Quantum-Theoretical Re-interpretation of Kinematic and Mechanical Relations"). Dover, New York

Heisenberg W (1926) Quantenmechanik. Naturwissenschaften 14:989–995

Heisenberg W (1958) Physics and philosophy. Harper and Row, New York
Hepp K (1999) Toward the demolition of a computational quantum brain. In: Blanchard P, Jadczyk A (eds) Quantum future. Springer, Berlin, pp 92–104
Hodgson D (1991 (hardcover) The mind matters. Consciousness and choice in a quantum world. Clarendon Press, Oxford
Koch C, Hepp K (2006) Quantum mechanics and higher brain functions: lessons from quantum computation and neurobiology. Nature 440:611–612
Lewis D (1979) Lucas against mechanism II. Can J Philos 9:373–376
Libet B, Wright EW Jr, Feinstein B, Pearl DK (1979) Subjective referral of the timing for a conscious sensory experience. Brain 102:192–224
Libet B, Wright EW, Feinstein B, Pearl DK (1992) Retroactive enhancement of a skin sensation by a delayed cortical stimulus in man: evidence for delay of a conscious sensory experience. Conscious Cogn 1:367–375
Lockwood M (1989) Mind, brain, and the quantum. The compound "I". Blackwell, Oxford
Lockwood M (1996) Many-minds interpretations of quantum mechanics. Br J Philos Sci 47:159–188
Maxwell N (1974) Can there be necessary connections between successive events? In: Swinburn R (ed) The justification of induction. Oxford University Press, London, pp 149–174. Am J Phys 40:1431–1435
McKemmish LK, Reimers JR, McKenzie RH, Mark AE, Hush NS (2009) Penrose-Hameroff orchestrated objective-reduction proposal for human consciousness is not biologically feasible. Phys Rev E 80:021912–021916
Nagel T (1976) Brain bisection and the unity of consciousness. In: Glover J (ed) The philosophy of mind. Oxford University Press, Oxford
Penrose R (1986) Gravity and state vector reduction. In: Penrose R, Isham CJ (eds) Quantum concepts in space and time. Oxford University Press, Oxford, pp 129–146
Penrose R (1989) The emperor's new mind. Oxford University Press, Oxford
Penrose R (1994) Shadows of the mind. Oxford University Press, Oxford
Penrose R (1996) On gravity's role in quantum state reduction. Gen Relativ Gravity 28:581–600
Penrose R (1997) Response by Roger Penrose. In: Longair M (ed) The large, the small and the human mind. Cambridge University Press, Cambridge
Pessa E, Vitiello G (2003) Quantum noise, entanglement and chaos in the quantum field theory of mind/brain states. Mind Matter 1:59–79
Planck M (1901) Ueber das Gesetz der Energieverteilung im Normalspectrum. Annalen der Physik [series 4] 4:553–563 (see also 556–57)
Ricciardi LM, Umezawa H (1967) Brain and physics of many-body problems. Kybernetik 4:44–48
Satari MV, Tuszynski JA, Zakula RB (1993) Kinklike excitations as an energy – transfer mechanism in microtubules. Phys Rev E 48:589–597
Schröedinger E (1926) An undulatory theory of the mechanics of atoms and molecules. Phys Rev 28:1049–1070
Schröedinger E (1935) Die gegenwärtige Situation in der Quantenmechanik (The present situation in quantum mechanics). Naturwissenschaften 23:807–812; 823–828; 844–849
Scott A (1996) On quantum theories of the mind. J Conscious Stud 3:484–491
Searle JR (1980) Minds, brains and programs. Behav Brain Sci 3:417–457
Searle JR (2007) What is language? Some preliminary remarks. In: Tsohatzdis S (ed) John Searle's philosophy of language. Cambridge University Press, New York
Smith CUM (2006) The 'hard problem' and the quantum physicists. Part 1: The first generation. Brain Cogn 61:181–188
Smith CUM (2009) The 'hard problem' and the quantum physicists. Part 2: Modern times. Brain Cogn 71:54–63
Squires E (1990) Conscious mind in the physical world. Adam Hilger, Bristol
Stapp HP (1993) A quantum theory of the mind-brain interface. In: Stapp HP (ed) Mind, matter, and quantum mechanics, 1st edn. Springer, Berlin, pp 145–172

Stapp HP (2001) Quantum theory and the role of mind in nature. Found Phys 31:1465–1499
Stapp HP (2007) Mindful universe: quantum mechanics and participating observer. Springer, New York
Stefan J (1879) Über die Beziehung zwischen der Wärmestrahlung und der Temperatur. In: Sitzungsberichte der mathematisch-naturwissenschaftlichen. Classe der kaiserlichen Akademie der Wissenschaften, Bd. 79 (Wien 1879), S: 391–428
Stuart CIJ, Takahashi Y, Umezawa H (1978) On the stability and non-local properties of memory. J Theor Biol 71:605–618
Stuart CIJ, Takahashi Y, Umezawa H (1979) Mixed system brain dynamics: neural memory as a macroscopic ordered state. Found Phys 9:301–327
Tegmark M (2000a) Importance of quantum decoherence in brain processes. Phys Rev E 61:4194–4206
Tegmark M (2000b) Why the brain is probably not a quantum computer. Inform Sci 128:155–179
Vitiello G (1995) Dissipation and memory capacity in the quantum brain model. Int J Mod Phys B 9:973–989
Vitiello G (2001) My double unveiled. Benjamin, Amsterdam
Von Neumann J (1955) Mathematical foundations of quantum mechanics. Princeton University Press, Princeton (German original *Die mathematischenGrundlagen der Quantenmechanik*. Springer, Berlin, 1932)
Wigner EP (1961) Remarks on the mind-body problem. In: Good IJ (ed) The scientist speculates: an anthology of partly baked ideas. Heinemann, London, pp 284–302
Wigner EP (1962) Remarks on the mind body question. In: Good IJ (ed) The scientist speculates: an anthology of partly-baked ideas. Basic Books, New York
Wigner EP (1967) Remarks on the mind-body question. In: Wigner EP (ed) Symmetries and reflections, scientific essays. Indiana University Press, Bloomington, pp 171–184
Wigner EP (1977) Physics and its relation to human knowledge. Hellenike Anthropostike Heaireia, Athens, pp 283–294. Reprinted in Mehra J (ed) (1995) Wigner's collected works, vol 6. Springer, Berlin, pp 584–593
Young T (1804) The Bakerian lecture. Experiments and calculations relative to physical optics. Philos Trans R Soc (Lond) 94 (Part I):1–16
Zeh HD (1970) On the interpretation of measurement in quantum theory. Found Phys 1:69–76
Zeh HD (1999) The arrow of time, 3rd edn. Springer, Berlin
Zurek WH (1991) Decoherence and the transition from quantum to classical. Phys Today 44:36–44

Chapter 16
Consciousness and Neuronal Microtubules: The Penrose-Hameroff Quantum Model in Retrospect

Eugenio Frixione

16.1 Introduction

If the number of diverse hypotheses, often absolutely at odds with each other, is a measure of collective ignorance about a subject in a given moment, surely the question of consciousness still holds one of the leading places in obscurity up to this day. Not only the issue usually referred to as "the hard problem"[1] with which most of this book is concerned, i.e., the human subjective experience of a self that integrates all sensory information, represents a persistent enigma. Indeed the whole set of phenomena included in the notion of consciousness remains, along with the puzzles of the origin of life and the state of the universe before the "big-bang," among the hardest and most contentious topics in today's science.

Yet few opinions would question that such a faculty is somehow an emergent property of the nervous system, dependent specifically on the brain. That is, nobody supposes anymore —as Aristotle and other precursors believed— that the center of awareness about the self and the world might reside in the heart or some other organ in the body. The cephalocentrists descendants of Alcmaeon of Croton, who lived in the fifth century BCE, won this aged discussion long ago.[2] In addition, we now know that the nervous system consists of a tissue built up by different types of cells, predominantly neurons and an assortment of so-called glial cells. Of these two main constituents, apparently only neurons show the highly complex connectivity required to function as an organized system for the efficient reception, conduction, transmission and integration of signals exchanged between diverse anatomical spots. In other words, as Santiago Ramón y Cajal demonstrated over a hundred

[1] The term "hard problem" is used here in its common conception as defined by Chalmers 1995.

[2] For a detailed historical review of this battle of concepts see Frampton 2008.

E. Frixione (✉)
Methodology and Theory of Science, and Department of Cell Biology,
Center for Research and Advanced Studies IPN (Cinvestav), Mexico City, Mexico
e-mail: frixione@cinvestav.mx

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_16,

years ago,[3] the nervous system is for the most part a network of individual neurons that are contiguous —but not continuous— with each other. Accordingly, any understanding of the faculty of consciousness will necessarily involve considerations about neurons and the ways in which they handle information.

Mainstream neuroscience holds that, whatever the specific mechanisms at work in each instance, short-term information handling by neurons is all a matter of electrical and chemical signals affecting primarily the plasma membrane that acts as the limiting surface in every cell. On the other hand, a marginal but resilient hypothesis matured in the mid to late 1990s insists that such a picture is only half of the story, and not the most important one. The standard neurophysiological mechanisms may be enough, its defenders claim, to execute the numerous neural responses involved in the managing of most bodily activities, such as carrying and distributing assorted sensory information from various origins to their appropriate destinations in the brain, as well as automatically processing part of such information and emitting suitable response commands so that fitting muscular and hormonal actions are accomplished on due time. This multi-level organization is obviously quite complex, yet essentially not too different from that mediating more basic reflexes like the common knee-jerk. In principle such a system represents a set of cybernetic arrangements capable of fine simultaneous regulation of a number of variables, through mechanisms akin to those widely at work in all of today's sophisticated computer-assisted systems.[4]

In the alternative view examined here, all of the above is hardly enough to account for consciousness and perhaps also for some of the other higher cognitive functions, like concept formation or reasoning and especially creative thinking, which lie far beyond the relatively simple associative memory involved in learning. It is argued that these refined faculties must include a much more elaborate managing of information, and that this probably takes place *inside* neurons, quite apart from —though clearly in relation and parallel to— those occurring at their plasma membranes. According to conclusions reached first separately and then further developed jointly by British physicist Roger Penrose and American anesthesiologist Stuart Hameroff, the more likely site where this intrinsically different kind of information processing might come about is within "microtubules," that is nearly ubiquitous intracellular structures otherwise established to play mainly mechanical roles in cells.

The "Penrose-Hameroff Orchestrated Objective Reduction (Orch OR) of quantum coherence in brain microtubules" is one of the quantum-based accounts of consciousness critically reviewed together in another chapter of this volume.[5] It represents a particularly complex subject in that it shares background of at least four different and quite distant research fields—advanced molecular cell biology, controversial quantum physics, juvenile cognitive neuropsychology, and infant quantum computing. Not too surprisingly, interdisciplinary discussion of the matter

---

[3] Ramón y Cajal 1899–1904; see also Loos 1967; Shepherd 1991; Lazar 2010.

[4] The classic seminal work on this approach to physiology was carried out by the mathematician Norbert Wiener in close collaboration with Arturo Rosenblueth and other leading neurophysiologists; see Wiener 1948 (a second revised edition appeared in 1961).

[5] Ward 2014.

has become especially difficult, with contenders frequently charging the opposing part with total misunderstanding.[6] The present review attempts to provide a mid-ground introductory explanation to the theory, against which the main debate issues are then briefly examined in retrospect.[7] An initial summary of neuronal cell biology is essential as a starting point.

16.2 Neuron Structure and the Cytoskeleton

Like any other animal cell, a neuron consists of three major components[8]: (1) the plasma membrane that, like a delicate skin, surrounds the entire cell and limits its extension in three-dimensional space; (2) the cytoplasm —previously called "protoplasm"— where all metabolic activity takes place and which is fully contained within the plasma membrane; and (3) a nucleus where most of the genetic information of the corresponding animal is conserved.

Unlike most other cells, however, neurons usually present a profusely ramified configuration (Fig. 16.1a) in which two main classes of offshoots can be distinguished: (a) multiple and progressively subdividing branches that project out from the region around the nucleus like boughs in a tree top, being hence called "dendrites" (from the Greek *dendron*, tree); (b) a single and comparatively long cylindrical fiber with an almost constant diameter called the "axon" (from the Latin *axis*), which may split into two or more runners at some points over its length though it typically branches out shortly before the end.

This apparently capricious morphology is a result of long functional optimization during evolution. Neurons adopt such shapes because branching greatly increases their surface-to-volume ratio, which in turn is favorable for the greatest possible connectivity among them, and at a convenient cost of metabolic maintenance too. A number of large-surfaced yet comparatively small-volume cells, so they can all be packed in a moderate little space, and each equipped with multiple appendages coursing in between many others, is the best arrangement for maximizing cell-to-cell contacts and interactions.

This morphological advantage, in turn, only becomes possible through the assistance of internal reinforcement provided by a framework of fibrous structures in each individual nerve cell (Fig. 16.1b, c). It is this "cytoskeleton" that permits a cell to grow in space with a peculiar shape, different from the roundish symmetry that it commonly has when just born by division of a mother cell. Moreover, apart from general structural support, the cytoskeleton also provides firm directional tracks along which proteins and intracellular organelles are actively transported by molecular motors, so as to furnish an adequate supply of components to every branch and branchlet from the usually distant metabolic center around the nucleus.

---

[6] See e.g., Grush and Churchland 1995; Penrose and Hameroff 1995; Tegmark 2000a, b; Hagan et al. 2002.

[7] A preliminary shorter version of this essay has been published (Frixione 2007).

[8] See Kandel et al. 1995, for general entry-level reference about neuron structure and function.

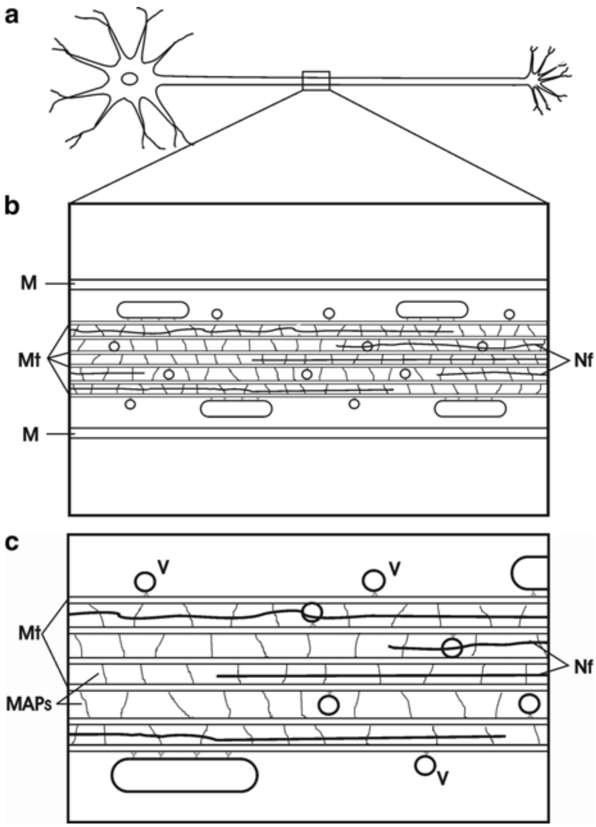


Fig. 16.1 General structure of a typical neuron. (**a**). The cell body (*left*) where the *nucleus* is located sends out two types of offshoots: (a) several relatively short branching extensions (dendrites), and (b) one often long cylindrical process (nerve fiber or axon) that usually also ramifies at the end (*right*). The dendrites and the cell body itself constitute the input area where nervous signals from other nerve cells are received and integrated. The product of such integration is eventually dispatched through the axon as an all-or-none electrical impulse known as "action potential." The outline of the drawing represents the *plasma membrane* (*M* in panel ***B***), in which all the bioelectric activity takes place. (**b**) and (**c**). Close-ups of a segment of the axon at two successively higher magnifications to illustrate part of the cytoskeleton, which in axons is constituted mainly by longitudinal *microtubules* (*Mt*) extensively cross-linked by several types of associated proteins (*MAPs*). This intracellular framework is further strengthened by parallel *neurofilaments* (*Nf*) that contribute to stabilize the ensemble. The cytoskeleton provides internal structural support for the whole cell, and offers firm scaffolding for the bidirectional transport of vesicles (*V*), mitochondria (elongated vesicles), and other cytoplasmic organelles that are carried along through mechanical work performed by molecular motors (Reproduced with permission from Frixione 2007)

As a consequence of this arrangement, the cytoskeleton is necessarily coextensive with the all-enveloping plasma membrane throughout the totality of a neuron or any other highly asymmetric cell. In other words, the plasma membrane always will be found accompanied by at least some tiny fibrous component of the cytoskeleton, which ensures structural stability and consistence of the cytoplasm even at the more remote reaches of any slender neuronal dendrite.

This last point is particularly important in our analysis because, as we shall see next, it is in one of the main components of the cytoskeleton —microtubules— where the quantum phenomena involved in consciousness are believed to occur according to Penrose, Hameroff and some other authors.[9]

16.3 Microtubules

All kinds of cells have at least one type of cytoskeletal structure, although it is of course in the more complex eukaryotic cells (those having a nucleus), which constitute the bodies of plants and animals, where the more developed cytoskeletons are found.[10] Animal cells, in particular, present cytoskeletons composed of three main linear structures: two types of filaments, each type with a characteristic thickness (actin filaments ~6 nm, neurofilaments ~10 nm), and long hollow cylinders measuring ~24 nm in external diameter called microtubules. It is the exquisite supramolecular construction of microtubule walls that has suggested a role far beyond mere structural support.

The microtubule wall is composed of 13 longitudinal "protofilaments" arranged in a circle so as to form a closed cylinder with a free central space of about 14 nm in width (Fig. 16.2). Each protofilament is in turn a linear polymer or straight chain

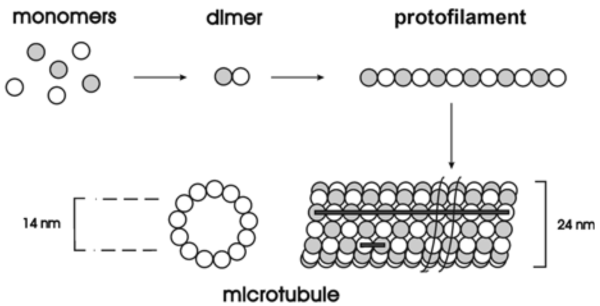


Fig. 16.2 Microtubule structure. Microtubules are hollow polymeric cylinders of ~24 nm in outer diameter with a ~14 nm central canal. They are constituted by two types of closely similar proteins known as α-tubulin and β-tubulin. These monomers associate as couples or *dimers* (one of them indicated by a *short black bar* on the microtubule wall), which in turn link up as linear longitudinal series called *protofilaments* (one of them indicated by a *long black bar* on the microtubule wall). A typical microtubule is formed by 13 of these protofilaments arranged in a *circle*, each slightly offset with regard to the adjacent ones so that the tubulin molecules also constitute a helicoidal array (indicated by *curved lines*) around the main axis (Adapted and reproduced with permission from Frixione 2007)

[9] Penrose 1994b; Hameroff and Penrose 1996.

[10] For a number of years cytoskeletons were thought to be an attribute of eukaryotic cells only, but they are now known to exist also in prokaryotes (cells without a defined nucleus) like bacteria. For a recent general review on the various fibrous structures constituting cytoskeletons see Frixione and Hernández 2011.

of dimers (couples) of two closely similar proteins, known as α-tubulin and β-tubulin, in alternating succession so that each couple constitutes a link in the chain. Because every protofilament is slightly offset longitudinally in relation to the two immediately adjacent ones, the microtubule wall is also actually a helicoidal lattice having an intrinsic structural bipolarity akin to that of screws and other helices. This polarity determines several properties. Thus, for example, the tubulin pairs can disassemble (depolymerize) from microtubule ends, or assemble (polymerize) again also at the ends, but both of these processes occur preferentially at one of the poles.

In neurons and their ramifications the microtubules are typically quite long, aligned with the main axis of the corresponding branch, and extensively interconnected by lateral linkages that include the microtubule-associated proteins (MAPs; Figs. 16.3 and 16.4). Similar links mediate also interactions with the plasma membrane and other elements of the cytoskeleton, like neurofilaments and actin filaments.

16.4 Microtubules Involved in Consciousness

As mentioned above, the precise interpretation of how can consciousness emerge from the neural apparatus varies among the different conventional, not quantum-based theories. All of these, however, assume without question that consciousness is at bottom a result of signals propagated along nerve cells in the form of electrical impulses, which may or may not give rise to new similar impulses in the neurons following immediately in the series. All processes at work, both during signal conduction along an axon, and upon signal transmission to a following target cell through a specialized intercellular junction called *synapse*, occur mainly at the plasma membrane on the surfaces of those cells. The model of consciousness based on microtubules admits a role for all of these conventional processes but insists that, while surely enough for actuating the body, these processes could hardly account for a faculty that belongs in an altogether different class.

This alternative hypothesis rests on three general premises.[11] First is the recognition that consciousness, construed primarily as capability for awareness, seems *not* to be a singularity of humans and other higher animals. On the contrary, just like many other products of biological evolution, "the *fabric* of consciousness may be present within all eukaryotic cells" —observes Hameroff[12]— as a continuous gradation in the living world, from rudimentary exploring behavior in unicellular

[11] Most of the extensive theoretical work provided as background for the basic proposal, going back to Charles Darwin's conviction that even consciousness must be a product of evolution, can be found in Hameroff and Watt 1982; Penrose 1989, pp. 411–413; Hameroff 1994; Penrose 1994a, pp. 41–61; Penrose 1994b; Hameroff and Penrose 1996; Hameroff 1998a, b; Penrose 2001. For a talk by Penrose on this subject log on to http://www.youtube.com/watch?v=f477FnTe1M0. Hameroff's two-part defense of their theory can be watched by logging to http://www.youtube.com/watch?v=ZAVQjMf2fEQ and http://www.youtube.com/watch?v=ed9nZXrOaMk

[12] Hameroff 1994.

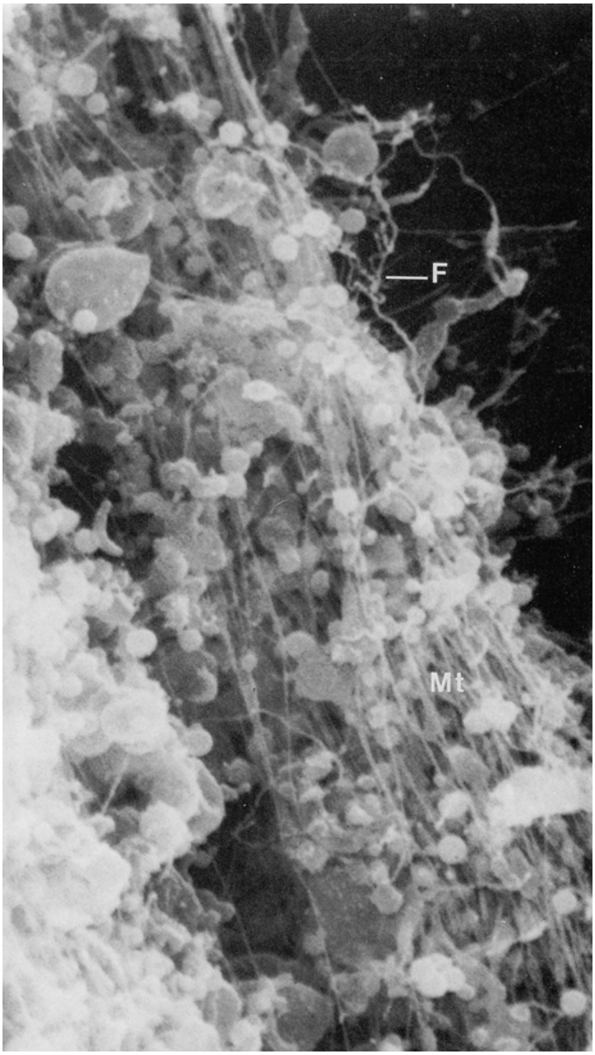


Fig. 16.3 Scanning electron micrograph of a column of microtubules (*Mt*) in a crayfish retina photoreceptor cell, in which numerous granules and other organelles remain attached to the long continuous microtubules after the plasma membrane was intentionally stripped off by a mild tissue-disrupting preparation procedure. Wavy filaments (*F*, probably neurofilaments) are also associated with both microtubules and organelles (Reproduced with permission from Frixione 1983; ©1983 Rockefeller University Press. Originally published in *J. Cell Biol.* 96:1258–1265)

organisms up to the most sophisticated levels in human *understanding*. Thus, for example, it can be argued that there is a primitive consciousness in relatively simple organisms and even in free-living individual cells, like protozoans, which are capable of finding out their way through obstacles in their path without the benefit of a single synapse. According to Penrose, this basic capacity apparently evolved to eventually

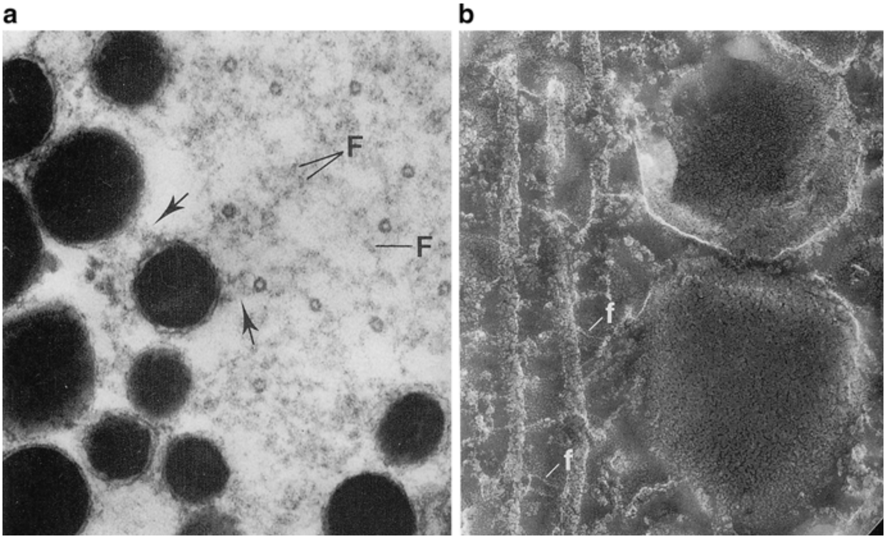


Fig. 16.4 High-magnification transmission electron micrographs of the cytoplasm of crayfish retina photoreceptor cells, like that shown in Fig. 16.3, after preparation of intact retinas by two different procedures. *Left*. A transverse thin section shows microtubules (*small gray rings*) and granules (*large dark masses*), all extensively cross-linked by a pervasive filamentous material (*F*) in which microtubule-associated proteins (MAPs) must be included. *Arrowheads* indicate granule-to-granule, as well as granule-to-microtubule fibrous linkages. *Right*. Longitudinal view of a similar cell region in a freeze-fractured and etched specimen, showing numerous connections between microtubules (*vertical rods*) and granules (*large spheroids*), including very thin filaments (*f*) that possibly correspond to MAPs (Reproduced with permission from Frixione 1983; ©1983 Rockefeller University Press. Originally published in *J. Cell Biol*. 96:1258–1265)

reach its most refined known expression in mathematical intuition and artistic inspiration.

A second premise in this model of consciousness is that, at the cellular level, the capacity for being aware seems to be intimately related to the cytoskeleton. In free-living unicellular organisms like the protozoans just mentioned, for example, the presence of nutrients or noxious substances in the environment is detected directly through receptor proteins located at the plasma membrane, and this in turn activates either approaching or retreating patterns of locomotion that are carried out by elements of the cytoskeleton. In the case of ciliates like paramecia, those elements are microtubules organized in a special array called axoneme, the actual motile apparatus within each cilium. Since the repetitive presentation of stimuli often leads to a diminution of those behavioral responses, it may be admitted that there is some habituation of the organism to such stimuli, and this implies the existence of something like a primitive "memory."

Now, there is no reason to suppose that the link between perceptive processes taking place at the plasma membrane and certain cytoskeleton-related responses is restricted to unicellular organisms. On the contrary, again, this may be a general property that only becomes particularly well developed in mammalian nerve cells. There is increasing evidence, for example, that the actin-filament component of the cytoskeleton serves an important role in the plasticity of the central nervous system associated with memory and learning, not only through its active involvement in relevant morphological changes of the synaptic junctions but also by directly modulating the efficacy of the neural transmission itself.[13] As regards specifically the microtubules, these have been shown to invade and thereby promote the growth of brain synaptic junctions in response to intense neural activity, thus also contributing to long-lasting ultrastructural changes related to learning.[14]

Finally, the third premise is that a micro-anatomical substrate with the right dimensions and a high degree of intrinsic organization, like the cytoskeleton and especially the system of microtubules, offers characteristics that are quite appropriate for quantum processing of information in the nervous system. The microtubules would function here as devices for a longitudinal integration of transient information, which then might be locally modulated and transversely communicated from one microtubule to others through the MAPs and other lateral linkages interconnecting them (see Figs. 16.1, 16.2 and 16.4). The extensive intracellular network constituted by microtubules and MAPs would thus represent the lower level in a hierarchy of functional webs across the nervous system, where neuronal networks of increasing orders of complexity represent the upper levels.

16.5 Microtubules as Capable Computing Devices

The input of information in this multi-level processing apparatus could take place in the following way. Electrical excitation at the plasma membrane of a neuron would induce, either directly by the currents flowing through activated ion channels, or indirectly through the resulting fluctuations in the levels of intracellular free calcium (Ca^{2+}), a discrete reversible change in the three-dimensional conformation of some tubulin dimers on nearby microtubules. Such reversible conformational change, common in many proteins —like the ion channels themselves—, would make a significant difference among microtubules: at any given moment those microtubules affected by the electrically or Ca^{2+}-mediated modification would have a fraction of their tubulin dimers switched from a certain configuration A (resting or "off") to another conformation B (excited or "on"; Fig. 16.5).

Should this change in turn induce or facilitate identical transitions in tubulin dimers situated in adjacent protofilaments of a microtubule wall —in a way analogous to how the change in transmembrane voltage by activation of sodium-ion channels facilitates the activation of subsequent sodium channels—, the number of "excited" tubulin dimers would likely tend to increase. Given the peculiar molecular arrangement in microtubule walls, where every subunit is at once a member of a straight

[13] Cingolani and Goda 2008.

[14] Merriam et al. 2011.

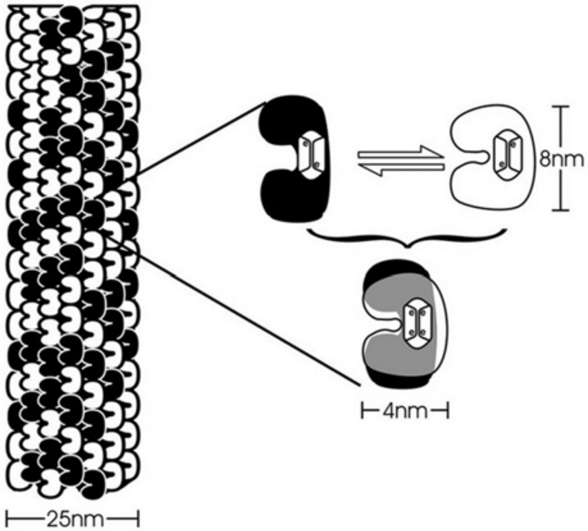


Fig. 16.5 Schematic illustration of conformational changes of tubulin dimers in microtubules, according to the "Orchestrated Objective Reduction" model of consciousness (Fig. 2 in Penrose and Hameroff 2011, reproduced here with permission). The tubulin dimers can either merely change in conformation between two alternative states (illustrated as *black* and *white* silhouettes on the *top right area*), or they also can enter into a quantum superposition of both states (combined silhouette *below*). In the first case the microtubule (*left*) can perform as an efficient switching matrix for classical digital computing, while in the second instance powerful quantum computing becomes possible

series (protofilament) *and* of a helicoidal series of units (see Fig. 16.2 above), each tubulin dimer constitutes a possible straight-ahead or sideward switch for the propagation of molecular excitation. Consequently, the microtubule walls would behave as "switching matrices" where information is represented by specific patterns of tubulins in either "on" or "off" states. Moreover, these patterns might advance along microtubules and rearrange under the effects of successive electrical events in the plasma membrane.[15]

In principle, distinct distributions of tubulin dimers in two different conformational states within the intracellular population of microtubules constitutes binary information (where each tubulin dimer in a given state represents one "bit"), which would make the cytoskeleton of every neuron a diminutive digital computer with potential for an incredibly high performance. The information processing capacity of the human brain with a theoretical configuration of this model —considering

[15] Hameroff and Watt 1982; Hameroff et al. 1986, 2002. For a brief background and the current status of these hypothetical mechanisms, which now have expanded to include molecular encoding of memory through phosphorylation of calcium-calmodulin II in microtubule lattices, see Hameroff et al. 2010; Craddock et al. 2012.

average density of microtubules per neuron, fraction of the brain constituted by neurons, and average rate of neuronal firing— has been calculated as of approximately 10^{23}–10^{26} bits per second.[16] By comparison, the figure estimated for conventional mechanisms based solely on electrical impulses and synapses is at most about 10^{16} bits per second.[17]

Nevertheless, mere computing capacity is not enough for explaining the emergent character of consciousness, which is essentially non-computable in the classical sense according to the proponents of the microtubule-based hypothesis. That is, the simple linking of algorithmic operations, as fast and complex as that might be, will never suffice for the system executing them to become conscious of itself. It is irrelevant whether such operations are electrical impulses conducted by nerve fibers and chemically relayed at synapses, or patterns of bimodal tubulin states moving along the walls of microtubules in their axoplasms, or electrons passing through artificial integrated circuits. Multiple intrinsically independent signals exchanged among different nodes within a switching network can perhaps produce a sophisticated resultant output, but cannot by themselves give rise to the sort of unified understanding that is evident in a dog or any other of the higher mammals. Neither totally deterministic nor known random simulations, in any of the varieties of each usually programmed to run in digital computers, will ever allow them to make *free individual decisions* such as a dog's in choosing to go left or right, or to develop an idiosyncratic taste for certain sequences of sounds.

The partly organized, partly indeterminate character of these latter phenomena, typical of mental processing, escapes the rigid computing-like mechanical explanations offered by classical psychophysiology.[18] Rather, the structured but non-mechanical quality of many cognitive processes seems akin to the uncertain nature characteristic of quantum mechanics. Hence, the notion that the phenomenon of consciousness might have a quantum-related origin has risen in a recurrent manner over the last 50 years,[19] stemming mainly from broad analogies between the mind-brain duality, on one hand, and the wave-particle double aspect of subatomic constituents, on the other. The proposal that microtubules are ideal candidates for the occurrence of such quantum processes, therefore, is just one relatively recent instance of a wider tradition that has kept emerging for several decades already. The filigree-like architecture of microtubules is supposed to undergird, apart from the switching-matrix binary computation mentioned above, a non-robotic and immensely more powerful kind of information processing, i.e., quantum computing, that allegedly can account for true consciousness. The following paragraphs attempt to provide a rough idea of how brain microtubules might support the required quantum computing.[20]

---

[16] Hameroff 1994; Penrose and Hameroff 2011.

[17] Moravec 1990, p. 61; see also discussion in Kurzweil 2005, pp. 122–126.

[18] Hameroff 1998b.

[19] See Smith 2006, 2009; also chapter by Ward in this volume.

[20] See Kurzweil 2005, pp. 120–121, for a compact explanation of quantum computing and its potential, and Litt et al. 2006, for a discussion in relation to consciousness.

16.6 Coherent Quantum States in Microtubules

The hypothesis supposes that the above-mentioned conformational change of tubulin dimers, from a resting or “off” state to an excited or “on” state, involves translocations of electrons in one or more of their constitutive atoms. In terms of atom physics tubulin “excitation” would consist of, or would be accompanied by, quantum transitions of those electrons to a different energy level. Since the conformational change would be identical in every “excited” tubulin dimer, at a given moment the tubulin dimers constituting the whole population of microtubules in an active neuron would be distributed in two different quantum states, each type corresponding to a specific wave function. All of the polymerized tubulin dimers in each of these two states would be found at the same energy level, which means that those molecules would be all quantum coherent.

Now, quantum transitions are not as simple as transitions in the common macroscopic world ruled by classic physics. In the latter, alternations of a system between two possible states may be equaled to flipping a switch between an “on” and an “off” positions, that is, the positions are actually exchanged as mentioned above. Thus the system is always found in either one or the other position, or it may even be in an intermediate position that is neither one of the two basic possibilities but still clearly defined as being midway.

Things are different in the submicroscopic world ruled by quantum mechanics,[21] where two (or more) possible quantum states of a system may co-exist in a *superposition* state, rather than being exchanged or intermediate. In fact, all the possible quantum states exist simultaneously in an indeterminate and even *entangled* condition. This ambiguous or rather “multiguous” state persists until the system is “observed” (e.g., measured); only then the ambiguity reduces into one of its multiple possibilities.[22] An informal comparison for intuitively grasping the notion of this apparently weird phenomenon may be found perhaps in those ambivalent images (optical illusions) that have two alternative interpretations, depending solely upon the viewer’s mental attitude. The drawing is actually indeterminate but acquires a definite significance when an observer looks at it in a given moment.

Tubulin dimers constituting a microtubule may thus conceivably be found, apart from in the resting (“off”) or in the excited (“on”) states discussed above, in a third state of quantum superposition, i.e., they would be coherent in this third state with its own wave function (Fig. 16.5). In terms of computing science, instead of representing simply one “bit,” each of these molecules can be regarded as constituting one quantum-bit or “qu-bit” of information.[23] Yet this is an inherently unstable con-

[21] For an accessible compact general introduction to quantum mechanics see Mielnik and Rosas-Ortiz 2009. Sections of especial interest for the present purpose are found in items 17–19, and 23.

[22] An early postulation of this intriguing conclusion was introduced by London and Bauer 1939 (see also English translation in Wheeler and Zurek 1983). For later formulations of the same idea see Wigner 1962, and Penrose 1994b. Criticism may be found in Bell 1987, and Hepp 1998.

[23] For Hameroff’s spoken explanation about qu-bits in microtubules log to http://www.youtube.com/watch?v=LXFFbxoHp3s

dition because unlike macro- and mesoscopic coherent systems, i.e., many-particle assemblies such as laser beams that can be robust enough as for perforating hard metal, quantum coherence of electrons in atoms is extremely sensitive to disturbance by interaction with the surrounding material medium. The motions of nearby particles impinging upon an electron system in a coherent quantum-superposition state as a result of thermal agitation, especially in liquid aqueous media, as well as interference of foreign wave functions, rapidly induce reduction of the system to a non-superposition state and thus decoherence. Hence, the development and maintenance of coherence in a quantum-superposition state is in principle quite unlikely, at least for any physiologically significant times, under the warm temperatures and thoroughly wet conditions in which living matter exists. Such coherence in a quantum-superposition state, if developed at all in common biological systems, should be extremely short lived. Decoherence timescales of ~10^{-13}–10^{-20} s, far too brief for any physiological process, have been calculated to occur in brain tissue.[24]

It can be argued, however, that the peculiar structure of microtubules and their teeming assemblies in neurons determine conditions that are quite different from those prevailing in the cytoplasm of most other cells. Given the arrangement of tubulin dimers in the wall of a microtubule, it may be inferred that both the outer and inner surfaces of the hollow cylinder offer physical properties that foster the adsorption of relatively large layers of water molecules. This water, virtually immobilized by electrostatic interaction of the dipole in every water molecule with fixed charges exposed on the adjacent tubulin dimers, is not in liquid state like the bulk water found in most of the cytoplasm. In fact it becomes structured water, analogous to ice, probably wrapping and filling every microtubule. The vicinity of other microtubules and cytoskeletal elements, like neurofilaments and particularly dense networks of crisscrossing actin-filaments constituting a gel, should also contribute to local water immobilization and overall noise reduction.

Accordingly, the immediate environment of a microtubule should not be wet and noisy, but relatively firm and therefore comparatively quiet. These conditions may be expected to be even more shielding and isolating in the narrow interior of a microtubule, which is likely occupied by a rod of highly structured water suited to act as a waveguide for an efficient longitudinal transmission of quantum energy. It can be theorized[25] that these qualities are favorable for the occurrence of optic phenomena known as "superradiance" and "self-induced transparency," such as those found in laser beams. Hence, again theoretically, every microtubule would offer a peculiar set of properties favorable for "tuning-up" (entangle) in a coherent quantum-superposition state tubulin dimers distributed along hundreds of microns or more over its own wall. Quantum-superposition coherence at such macroscopic scales is sometimes called a "Bose-Einstein condensate," and might be not so unusual in biological systems.[26] Evidence for macroscopic quantum coherence

[24] Tegmark 2000a, b.

[25] Jibu et al. 1994.

[26] Fröhlich 1968.

mediating wavelike energy transfer has been obtained recently for the highly organized protein complex that executes photosynthesis.[27]

It is hypothesized, therefore, that numerous "off-on" alternations in tubulin dimer states over the wall of a microtubule at some point give rise to a coherent superposition state, which eventually might extend throughout the length of the cylindrical polymer. Now, the uniform geometry of every microtubule, the quasi-crystalline linear-*and*-helicoidal arrangement of tubulin dimers making up its wall, and especially the alignment in parallel with many other microtubules, which for the most part (if not all) are extensively cross-bridged by regularly spaced MAPs and other filamentous linkages, all contribute to a multi-level orderly system of biological micro-organization that is especially appropriate for long-range interactions between molecules. It is then conceivable that the MAPs transversely linking neighboring microtubules at roughly regular intervals might help to spread the coherent quantum-superposition state of tubulin dimers across the entire cytoskeleton of a neuron.[28]

Furthermore, quantum coherence is suspected[29] to spread also from neuron to neuron by a tunneling effect through gap-junctions, namely passages directly communicating the cytoplasms of contiguous cells. Simultaneous or concurrent electrical activity in synergic neurons, or in those belonging to functionally related nervous pathways, as well as in specific neural networks, might also contribute to increasing quantum coherence among polymerized tubulin molecules located in different regions of the central nervous system. Accordingly, in theory at least, a physical substratum might exist for the coupling of quantum dynamics in huge numbers of polymerized tubulin dimers throughout vast regions of an animal brain. But, what has all of this to do with consciousness?

16.7 Periodic Self-Reductions in Coherence of Quantum Superposition

The previous section discusses how coherent quantum-superposition of tubulin dimers supposedly occurs in neuronal microtubules, even under the superposition-adverse physical conditions (mainly high temperature and wetness) in which living matter normally exists. Tubulin dimers are believed to be gradually recruited in the quantum-superposition state as coherence propagates, first over a few microtubules and then progressively throughout a number of neuronal cytoskeletons in the system. Still, the argument goes, this quantum-superposition state is a transient condition that increases and spreads over time up to a limit, after which coherence is necessarily lost. In sum, the process continues to build up so that coherence of quantum superposition becomes increasingly difficult to sustain, until the whole collective

[27] Engel et al. 2007.

[28] See Woolf 2006, for a discussion of how acetylcholine-dependent phosphorylation of MAP2 might regulate the quantum-computing ability of microtubules.

[29] Woolf and Hameroff 2001.

superposition state undergoes a sudden collapse or reduction (decoherence). A theoretical example suggests that tubulin dimers could remain coherent in the quantum superposition state for as long as 25 ms within dendrites of pyramidal cells interconnected by gap-junctions in the visual cortex.[30] A series of these events involving progressively higher numbers of tubulin dimers would constitute a "visual epoch" lasting up to 500 ms.

This, in turn, begs the question of what determines such a reduction, since the peculiar local conditions within neurons are indeed so uniquely favorable for quantum coherence despite the adverse factors, as discussed above. Penrose and Hameroff propose that, as the system exceeds a maximum limit of quantum-coherent superposition, a local disruption in space-time geometry occurs under the effect of a property called "quantum gravity."[31] That breakup is accompanied by an instantaneous loss of all quantum-coherent superposition despite the absence of an external observer or subject as inducer, being therefore a self-promoted or "*Objective* Reduction" (OR). A crude metaphor may be helpful to intuitively form a mental picture of such a catastrophic event.[32]

Imagine a growing house of playing cards in which each card can be in only one of three possible positions—laying flat horizontally anywhere on the site with the figure either facing down (1) or facing up (2), or standing vertically on one edge upon other cards and supporting others somewhere within the precariously balanced construction (3). Cards laying flat horizontally can be made to spell information if, for example, a binary code of face-up and face-down is established and they are then tallied accordingly in linear progression. Moreover, the information handling of this system can be dynamic if the sequences of face-up and face-down cards change over time, like tubulin dimers alternate between "off" and "on" states in microtubules according to the Hameroff and Penrose's model. In contrast, standing cards cannot be counted in this way because they always show both faces to plain view, although just for a necessarily limited period. As more cards are successively piled up, one over another in increasing numbers, the stability of the structure gradually diminishes until a certain size is reached when the construction collapses under gravity.

This crash is not a purely random event, since it is clearly related to the amount of *mass* (number of cards) in an intrinsically unstable though nevertheless *coherent* (standing) *state*, to the *space* involved (size of the house), and to the *time* elapsed since the beginning of the construction; the higher each of these variables becomes, the higher the probability of collapse. Still, although maybe approximately foreseeable, the outcome is not predictable by any algorithmic procedure of classic computation; at every try, different numbers of piled cards, house sizes, and standing times will be recorded before the breaking point is reached. Nor will it be possible to anticipate in which direction the building will preferentially lean as it begins to come down, or

[30] Ibid.

[31] Penrose 1996; Penrose and Hameroff 2011, Section 6.

[32] The author assumes all responsibility for this and the following analogy below, introduced here solely for facilitating the explanation of these concepts to an heterogeneous readership.

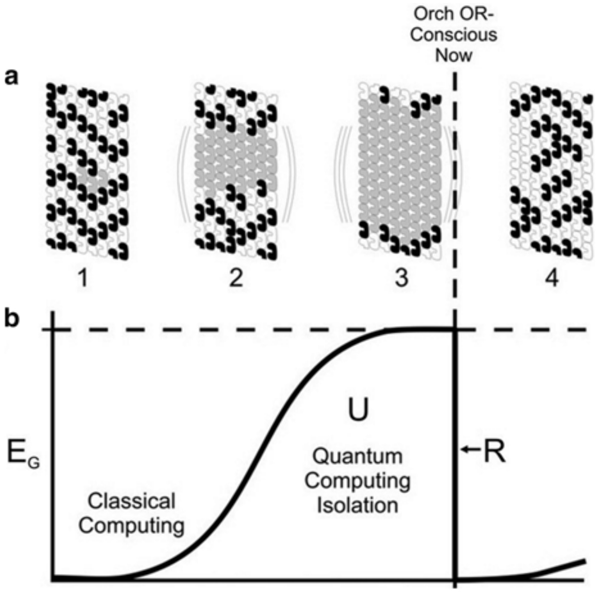


Fig. 16.6 Consciousness in relation to coherence of tubulin dimers in the quantum-superposition state over the walls of neuronal microtubules, according to the "Orchestrated Objective Reduction" (Orch OR) model of consciousness (Fig. 5 in Penrose and Hameroff 2011, partial reproduction with permission). (**a**). Switching of tubulin dimers between a resting or "off" state (*white particles*) and an excited or "on" state (*black particles*) form changing patterns that mediate classical digital computing, and leads many tubulin dimers to progressively enter into a coherent quantum-superposition state (*gray particles*) compatible with quantum computing. (**b**). This recruiting of tubulin dimers in a coherent quantum-superposition state (*U*) continues up to a limit of "gravitational self-energy" (E_G), at which point Orchestrated Objective Reduction (*R*), i.e., instantaneous collapse of coherence, occurs along with a brief instant of conscious experience. Then a new "epoch" of building quantum-superposition coherence begins among the population of tubulin dimers

how many cards will still be perhaps left standing at the base of the house once the rest are all lying scattered around, or the positions (face-up, face-down, and orientation) of the latter. That is, there is a fundamental uncertainty inherent in this simple system; hence it is not computable. At best one can obtain probabilistic numbers of the variables involved after a few tests.

Analogously, according to the Penrose-Hameroff's hypothesis, the coherence of tubulin dimers in the quantum-superposition state over the walls of neuronal microtubules will collapse in a non-computable way under the action of quantum gravity (Fig. 16.6a), as a function of the number of dimers in the quantum-superposition state, the space over which the coherence extends, and time. And just as the period during which the house of cards remains standing and growing will shorten by the occurrence of unfavorable external factors, such as wind or vibrations, so the coherence of tubulin dimers in the quantum-superposition state will collapse depending upon adverse conditions in the environment, especially thermal noise in liquid aqueous systems like those prevailing within living cells.

16.8 The Emergence of Consciousness

How such a self-collapse or "Objective Reduction" (OR) in quantum coherence might be related to consciousness could be explained with still another —in this case more physiological— comparison. Sensory information carried from the sense organs to the brain is encoded as electric nerve impulses that travel along afferent nerve fibers, each of these impulses consisting of a transient sudden local collapse (which may include even a swift partial reversal) of the voltage difference normally existing across the plasma membrane while a nerve fiber is at rest. This phenomenon is a unitary all-or-none response that does not vary in magnitude regardless the intensity of the stimulus. Yet the crucially important information about stimulus intensity is effectively communicated to the central nervous system through time encoding: that is, stronger stimuli will produce more frequent electric impulses than weaker stimuli. Thus a largely "subjective" attribute like intensity is translated into a purely "objective" quantity, which the system can interpret accordingly. Whereas a single electric impulse in a sensory fiber is hardly detectable at all as a sensation, a number of impulses in rapid succession are perceived as a definite stimulus of certain intensity. The significance of the signal will be higher, of course, as the number of similar sensory fibers conducting electric impulses from the same body part is increased.

Likewise, the collapse of quantum coherence on the wall of a single microtubule, that is, just one minimal "element of *proto-consciousness*" (Fig. 16.6b), could be regarded as probably undetectable in itself. Nevertheless, simultaneous or closely timed collapses of coherence in neighboring microtubules might compose a definite signal. If this process occurs at several places in the central nervous system at the same time —e.g., in visual, auditory and olfaction areas—, spatial dimensions are added to the temporal one so that a multi-polar spatio-temporal unification would arise. This, which perhaps may be likened to the significance that builds up by the integration of clusters of sounds with different frequencies and timbres, all coming from separately located instruments in an orchestra, has been called an "*Orchestrated* Objective Reduction" (Orch OR) in quantum coherence.[33]

Now, if Objective Reduction in quantum coherence in one microtubule constitutes the elementary unit of conscious experience, just like an electric impulse (a transient collapse in membrane voltage) constitutes the elementary unit of a sensory experience, then an Orch OR in quantum coherence could theoretically amount to an instant of awareness. This last phenomenon has been experimentally determined to occur at about every half a second (~500 ms),[34] the average time it would presumably take to build up pre-conscious coherent quantum-superposition of tubulin dimers to the point of another Orch OR. A succession of Orch OR's would then

[33] One of the editors of this volume noted that René Descartes would have been disappointed to learn that, according to this model, the immaterial faculty of thought, his famous *res cogitans*, is after all distributed in space just like the *res extensa* or mere matter.

[34] Libet et al. 1979.

produce the subjective impression of a continuous stream of consciousness, just like trains of electric impulses carried to the brain through fibers of the auditory nerve are perceived as continuous sound. Synchronous binding of intrinsically separate processes into an apparently unified spectrum of instantaneous experience would be thus achieved.

It is to be noted here that not a majority or even a large number of neurons would need to be involved in the above process to generate consciousness, like not every instrument in an orchestra nor most of them are required to play at the same time in order to produce music. Basic calculations[35] indicate that quantum coherence of just one percent of the tubulin dimers present in 20,000 neurons, built for some 25 ms, would be required to reach an Orch OR. Since the focus of consciousness varies over time, some kind of navigation system is necessary to integrate and prioritize the processing of merely internal information (thought, memory, daydreaming, and so on) with the shifting variety of perceptions received from the body and the outside world. In order to fulfill this requirement, a dynamic "conscious pilot" has been recently suggested to steer quantum coherence through the central nervous system.[36] The vehicle proposed for this overall integrator is the gamma band (30–100 Hz) of electroencephalographic activity, commonly believed to derive from a synergic interaction of neuronal groups. Such interaction is apparently mediated by bridges of cytoplasmic continuity (the gap junctions mentioned above) being momentarily formed between cortical dendrites, so that temporal networks arise which could conceivably allow synchronized activity to move from place to place along particular paths. Thus "a mobile gamma-synchronized dendritic web" could travel as a spatiotemporal envelope that integrates quantum phenomena in different regions of the brain.

Anything capable of preventing the polymerized tubulin dimers from effecting the hypothetical conformational change that leads to the quantum superposition state is bound to have profound consequences on consciousness, of course. This might be the case, for example, with several gases that share an unspecific high affinity for hydrophobic regions in proteins, though they are otherwise chemically unrelated — like ether, chloroform, trichloroethane, halothane, and some derivatives of flurane, i.e., the general anesthetics commonly used to induce loss of consciousness. Hameroff speculates[37] that these substances interact with hydrophobic pockets involved in the regulation of conformational transitions of neuronal proteins involved in consciousness, including those found in microtubules and in the gap junctions that mediate synchrony of the electroencephalographic gamma band.

As it can be deduced from the previous summary, the Penrose-Hameroff Orch OR model of consciousness seems to have a good deal of the underpinnings required to resist critical scrutiny. Let us now take a brief look at how has it been received.

---

[35] Penrose and Hameroff 2011; see also Hameroff and Penrose 1996, and Hagan et al. 2002.

[36] Hameroff 2007, 2010; Ebner and Hameroff 2011.

[37] Hameroff et al. 2002; Hameroff 2006.

16.9 Reactions and Contrarreactions

Penrose and Hameroff have faced severe criticism from the very inception of their joint theory on consciousness. Although the initial almost mocking reaction[38] was addressed mainly to Penrose's then recently published book *Shadows of the Mind*,[39] Hameroff was mentioned as well so they co-authored an immediate point-by-point reply in the same journal.[40] Apart from the quite debatable issue of whether human thought is ultimately non-algorithmic (non-computable), only because according to Gödel's incompleteness theorems there is no consistent set of axioms that is valid for all mathematical systems,[41] Penrose's suggestions were deemed entirely speculative, without evidences for a series of strong suppositions, and with no valid connection to either consciousness or microtubules. As regards the latter, Grush and Churchland pointed out that no impairment of consciousness is known to occur in gout patients medicated with colchicine, a drug that induces microtubule depolymerization. The explanation for this inconsistency, argued correctly Penrose and Hameroff in their reply, is that for all practical purposes colchicine does not cross the blood-brain barrier and thus it hardly reaches the microtubules in brain neurons. A further consideration is that, because of their extensive cross-linking and other local stabilizing factors, microtubules in neurons are comparatively resistant to the drug.

The Achilles heel of the Penrose-Hameroff's and other quantum theories of consciousness, however, has been explaining how quantum coherence can be built up and maintained for physiologically relevant intervals ($>10^{-3}$ s) across brain-size macroscopic scales, under the extremely unfavorable conditions of a wet and warm intracellular environment like that in mammalian neural tissue. As mentioned above, the harshness of molecular (thermal) agitation under such circumstances is immediately deleterious for even submicroscopic local quantum-coherent states. We shall recall[42] that this objection was thoroughly examined at the close of the century by Tegmark,[43] who calculated extremely brief quantum-decoherence rates ($\sim 10^{-13}$–10^{-20} s) that are absolutely incompatible with any quantum-coherence theory of consciousness in warm-blooded brains.

Once again the response was strong.[44] Tegmark's conclusions were found flawed on several counts, including confusion of the models being discussed and contradictions with observed facts. Moreover, revision of key calculations resulted in decoherence times compatible in principle with macroscopic quantum-superposition

---

[38] Grush and Churchland 1995.

[39] Penrose 1994a.

[40] Penrose and Hameroff 1995.

[41] For a thorough discussion on this matter see Litt et al. 2006; a compact examination of the point can be found in this volume (Ward 2014, Section 5).

[42] See Sect. 16.6.

[43] Tegmark 2000a, b. See also Hepp 1998; Koch and Hepp 2006.

[44] Hagan et al. 2002.

states in biological systems. Calculations of exceedingly short decoherence times were regarded as simplistic by independent authors also.[45] Subsequent examination of the controversy produced a mixed verdict,[46] where decoherence is attributed to interaction with the environment instead of to Penrose's "quantum gravity" factor, but with collapse times that keep open the possibility of the microtubules potential for information transfer and storage "at temperatures close to the human body."[47] A more recent analysis[48] of this issue cautiously concurs in that there are alternatives to explain quantum effects in brain neuronal microtubules, and an even more recent preliminary report[49] claims to have experimental evidence for quantum states in microtubules.

Problems have been raised also from the perspective of the basic biology of nerve cells. As pointed out by Smith,[50] the neuronal microtubules do not seem to be the quiet, undisturbed structures dedicated to classical and quantum computing envisaged in the Orch OR model. On the contrary, they are the constantly beaten tracks upon which molecular motors like kinesins and dynein concomitantly drag in opposite directions through the crowded axoplasm their bulky loads of vesicles, granules and organelles as large as mitochondria.[51]

Another direct critique leveled at the "Orchestrated Reduction" model of consciousness —in this case from the chemistry angle— is that (1) tubulin dimers cannot undergo conformational changes while they constitute microtubule walls, but only upon polymerization or depolymerization; and (2) that no heat radiation is detected from the tissue as it would be expected from the enormous amount of metabolic energy expenditure (GTP hydrolysis) for the simultaneous switching of conformation in huge numbers of dimers.[52] These authors are "clearly mistaken" in the first point, retorted Penrose and Hameroff,[53] because of a basic misunderstanding about the number of molecular regions (benzene rings) presumably engaged in tubulin switching between alternative states. The second point was granted, however, though not without stressing that such switching "need not actually involve significant conformational change," as Penrose and Hameroff's own usual but admittedly misleading diagrams illustrate, so that a high energy release (GTP hydrolysis) is not necessary for the conversion.

The Penrose-Hameroff model is vulnerable too, in general, to the entire list of objections so far raised against all paradigms of consciousness resulting from

[45] Davis 2004.

[46] Rosa and Faber 2004.

[47] Faber et al. 2006.

[48] Craddock et al. 2009; Craddock and Tuszynski 2010.

[49] Bandyopadhyay 2011.

[50] Smith 2009.

[51] Hollenbeck and Saxton 2005.

[52] McKemmish et al. 2009.

[53] Penrose and Hameroff 2011. Hameroff argues this point in a two-part video that can be watched at http://www.youtube.com/watch?v=ZAVQjMf2fEQ and http://www.youtube.com/watch?v=ed9nZXrOaMk

quantum processes in the central nervous system. Many of such objections are summarized and debated by Thagard and other proponents of neural or mental computation,[54] as opposed to quantum and ordinary digital computing. Because the dispute is largely focused on the theory held by Penrose and Hameroff, however, once again the latter author came out to its defense.[55] He admits that simple neurocomputation may have a role in mental phenomena, being thus compatible if not complementary with quantum computing according to the "Orchestrated Objective Reduction" model. On the other hand, though, he questions "whether neurocomputation alone can physiologically account for coherent gamma synchrony EEG, a candidate for the neural correlate of consciousness."

Finally, Penrose's contentions and implications have also been attacked from the camp of artificial intelligence enthusiasts.[56] Standard machine-based computation has already demonstrated its capability of mimicking many of the skills so far believed to be a monopoly of the human mind (like, notably, chess playing), it is argued. Thus, contrary to Penrose's claims, there is no intrinsic reason why this progress could not go on to include all the other mental abilities as well. Consciousness cannot represent a barrier in this trend, even if it results from non-standard quantum computing, because artificial quantum-computing systems are being developed already. Therefore, if higher cognitive processes are possible in biological systems, eventually they could be replicated in artificial ones as well. It is only a matter of time.

At bottom, the crux of the problem for a wider acceptance of the Penrose-Hameroff quantum theory of consciousness lies not so much in either physics (fast decoherence rates in warm wet milieus), physical-chemistry (energy involved in tubulin conformational changes), biology (known properties and functions of microtubules), or computing science (capabilities of quantum computing *per se* and the possibility of its occurrence in the brain). Nor is it the combination of all of these either, since as mentioned above they have all been confronted with arguments and data, disputable as these may be. Furthermore, the true applicable limits of quantum mechanics are still in the process of being determined, with quite a bit of it yet pending to be developed, so its role at the mesoscopic and macroscopic levels cannot be presently ruled out.

The unanswered question at the core of the problem is explaining how, even conceding the occurrence of successive cycles of coherence/decoherence in hypothetical quantum states of tubulin dimers in brain microtubules, *that by itself* might be perceived as conscious experience. In this sense, the difficulty is analogous to that of explaining just how a series of volleys of electrical activity in the visual cortex of the brain is *experienced* as "vision." In this respect Penrose and Hameroff leave the enigma of consciousness as enshrouded in mystery as ever, despite their ingenious

[54] Litt et al. 2006. See also Thagard's previous paper (2002) criticizing the limitations of ordinary computational models of mind for not taking into account the chemical complexity of nerve processes.

[55] Hameroff 2007.

[56] Kurzweil 2005, pp. 450–452.

proposal. As leading neuroscientist Christof Koch reportedly put it in a 1994 meeting on consciousness at the headquarters of the Penrose-Hameroff team in the University of Arizona: "Quantum mechanics is mysterious, and consciousness is mysterious. Q.E.D.: Quantum mechanics and consciousness must be related."[57]

16.10 Historical Precedents

The "Orchestrated Objective Reduction" theory of consciousness came as a totally unexpected (and unpleasant in some quarters) surprise for both neuroscientists and cell biologists, right at the end of a century full of discovery in both fields. Historically, however, placing the major significance of neural activity on intracellular linear structures, rather than on the plasma membrane of nerve cells, is only the latest chapter in a series that goes back for at least 300 years.[58]

In the early eighteenth century, upon the astonishing revelations made over the previous decades by the pioneers of microscopy, natural philosophers concluded that one of the secrets of living matter is its being structured as arrays of fibers at multiple dimensional levels. Thus Herman Boerhaave (1669–1738), the highly influential professor of medicine and chemistry at the prestigious University of Leiden, taught that the "animal spirit" mediating all nerve action, that is, "the most subtile Juice of any in the whole Body … being prepared and separated in the wonderful Fabric of the Cortex, is thence propell'd from every Point thro' these Tubuli …"[59] These "least vessels" or finest submicroscopic ducts in contemporary anatomical theory, explained later one of his prominent pupils, were "made up of simple fibres united together," and each "most simple fibre consists of very small parts adjoining to each other length-ways."[60] The actual structure of such hypothetical minimal "tubuli" was therefore conceived as polymeric both longitudinally and transversely, just like microtubules are built of protofilaments made up of aligned tubulin dimers.

Scientists of the Enlightenment did not have yet our current category of "cell" in their theoretical system, so they could not refer to those most slender "tubuli" as either intracellular or extracellular. A hundred years later, however, the new microscopists entered into a heated debate following Robert Remak's[61] finding of tiny fibrils *within* fresh nerve cells of the crayfish. The combative Max Schultze subsequently reported similar fibrils in ganglion cells of the torpedo fish and, in an era dazzled by the miracle of instantaneous long-distance communication via the telegraph wire, he maintained "the possibility of isolated conduction in these constituent fibrils."[62] This exciting idea was later picked up even by a young Sigmund

[57] Horgan 1997, p. 173.

[58] Frixione 2000.

[59] Boerhaave 1742–1746, vol. 2, §274, p. 285.

[60] Swieten (English translation), 1744, vol. I, Sect. 38, p. 98; and Sect. 21, p. 39.

[61] Remak 1843, 1844.

[62] Schultze 1870–1871, Vol. I, chap. 3, pp. 147–187 (English translation).

Freud who, in the aftermath of his own detailed study of crayfish neurons,[63] wondered whether the "fibrils of the nerve have the significance of isolated paths of conduction."[64]

At the turn of the twentieth century the electrical function of such intracellular threads was already tacitly admitted in many scientific circles, after both Stephan von Apáthy and Albrecht Bethe independently showed continuous fibril paths throughout nerve cells and fibers, using two different methods of histological preparation that produced exceptionally crisp staining.[65] This development called again into question the just recently victorious doctrine of the nervous system as built of discrete neurons, arguing instead in favor of the previously established continuous-reticulum scheme defended by Camillo Golgi and others. It took several years of hard work for Santiago Ramón y Cajal to become the undisputed leading master of neurofibril staining and study, and to dispel for good that new challenge to the universal acceptance of the true histological organization of the nervous system.[66]

Almost another full century passed for a new attempt at charging microscopic linear structures with nervous functions, if now in combination with the plasma membrane of the nerve cells, as Penrose and Hameroff presently claim. This by itself has contributed to a fresh new era of investigation on the cytoskeleton. In retrospect we can now say that microtubules have acquired a significance far beyond their traditional roles as relatively rigid structures capable of keeping the characteristic morphology of a cell, doubling also as tracks to support and guide the ceaseless multidirectional transport of intracellular organelles.

Apart from increasing evidence for the key participation of microtubules in the usual modes of information handling by neurons,[67] they have become a focus of attention for physicists and technologists.[68] Research in progress now deals with phenomena such as microtubules acting like biomolecular transistors for the amplification of electrical information,[69] and as pathways for intracellular ionic conduction,[70] while other approaches probe the stochastic resonance of tubulin dimers.[71] Irrespective of what the "Orchestrated Objective Reduction" model and other quantum-based accounts of brain function might hold for the future, they are already proving to be active promoters of new and potentially crucial knowledge at a most fundamental level.

63 Freud 1882.

64 See Frixione 2003.

65 Apáthy 1897; Bethe 1900.

66 Frixione 2009.

67 Chang et al. 2011; Merriam et al. 2011.

68 Jaeken 2007.

69 Priel et al. 2006.

70 Sataric et al. 2010; Craddock et al. 2010. For an interesting and apparently forgotten pioneering work in this line of research see Hejnowicz 1970.

71 Pizzi et al. 2010; Saha et al. 2012.

16.11 Concluding Remarks

The Penrose-Hameroff model of consciousness is a bold and interesting conjecture. Despite having been constantly assailed on various fronts, it keeps alive in the early 2010s after over 15 years of its first appearance, and still happily referring to "a connection between brain biomolecular processes and [the] fine-scale structure of the universe,"[72] which can hardly be denied about this or anything else. In Penrose's view, if there were indeed "a physical 'theory of everything' [it] should at least contain the seeds of an explanation of the phenomenon of consciousness."[73]

It surely counts among the better-known quantum-based theories of consciousness and is therefore unavoidable in reviews, even if often only for criticism or debate. Colorful animations of microtubules, while Stuart Hameroff explains, have been produced for popular television shows on scientific subjects.[74] Moreover, some of the theoretical principles have extended to include other relevant and more general aspects of cell life in which microtubules play prominent roles, like cell division (mitosis), cell differentiation, and some of the pathological sides of these processes, like cancer.[75]

Thus far the model still keeps strictly focused on its main subject —consciousness—, taking care of not straying too much into the subconscious and unconscious realms, other than providing the above cited account for general anesthesia. The theory also remains as pure as ever, that is, purely theoretical. Little if any direct experimental evidence is provided along with the wealth of data obtained from other sources. More to the point in connection with the present volume, the difficult issue of the *qualia* or raw ineffable attributes of sensory experience for individual humans, which lies at the very center of the "hard problem," is approached in terms of a "pan-protopsychism [so that] qualia are patterns in fundamental spacetime geometry accessed and selected by the Orch OR process."[76] Just how this selection is carried out is a point not discussed in this context, which instead goes on to compare the timings recorded in some Buddhist meditations. The reader is assured, however, that "If experiential qualia are qualities of spacetime, then Orch OR indeed begins to address the nature of conscious experience in a serious way."

The "Orchestrated Objective Reduction" model of consciousness shares most of the drawbacks, as well as many of the intriguing promises, of all quantum-based hypotheses that attempt to account for the higher faculties of the mammalian nervous system. But opinions on these topics could not diverge anymore. According to some views based on (or favorable to) leading quantum physics, the classical model of

[72] Ebner and Hameroff 2011; Penrose and Hameroff 2011.

[73] Hooft et al. 2005.

[74] Examples of repercussion in the media can be found in http://www.quantumconsciousness.org/media.html

[75] Hameroff 2004.

[76] Hameroff 2001.

neuroscience is but "a holdover from physical theories of an earlier era."[77] Nevertheless, such old-fashioned physical theories still continue to nourish traditional neurophysiology that, aloof and oblivious to quantum-based hypotheses of brain function, goes on explaining consciousness as related to simultaneity in neuronal firing derived from passive and active conduction at the level of cortical dendrites.[78]

At least part of the explanation for this schizoid character of current neuroscience may be found perhaps in the as yet tentative status of quantum computing itself in non-biological systems. According to some experts in this field, today's "quantum computers are still just toys in a test tube performing calculations a child can do" in her/his head.[79] On the other hand, alternative approaches are producing interesting though puzzling results. A recent report describes a novel method of *quasi-quantum* computing that has all the features of quantum computing (superposition, entanglement and collapse), even though it is executed using "conventional electronic devices in rather unconventional ways."[80]

As for the persistent problem of how to explain consciousness, perhaps the challenge may be compared to that posed for a very long time by the nerve impulse itself, believed to be mediated by the above mentioned "animal spirit" since at least Galen's teachings in the second century CE.[81] When the mechanism of transient voltage fluctuation across the plasma membrane was finally discovered,[82] following a century and a half of hard experimental work after the first realization that it was of an electrical nature in the early 1790s,[83] it became clear why all previous attempts were doomed to fail. Neither all the necessary empirical knowledge about the system, nor the sets of mind required to understand that knowledge, were then available to even begin to approach the issue.

This may be also the situation with consciousness and the "hard problem" in the cognitive sciences. Here too, as Schrödinger put it almost 70 years ago, "living matter, while not eluding the 'laws of physics' as established up to date, is likely to involve 'other laws of physics' hitherto unknown, which, however, once they have been revealed, will form just as integral a part of this science as the former."[84]

Presently we really do not even know how far advanced is the progress of research in the field of consciousness. Nevertheless, should it turn out in the future that quantum computing is indeed involved in this striking phenomenon, microtubules may still count among the most likely candidates for coherence and decoherence of quantum-superposition states, at least in the mammalian central nervous system.

---

[77] Schwartz et al. 2005.

[78] See a classic but still valid example in Llinás et al. 1998; see also relevant discussions in Koch and Hepp 2006; Ward 2014.

[79] Kendon et al. 2010.

[80] Haikonen 2010.

[81] See recent book on this subject by Smith et al. 2012.

[82] Hodgkin and Huxley 1952.

[83] Galvani 1791.

[84] Schrödinger 1944/1994, p. 495.

Acknowledgments The author wishes to dedicate this essay to the memory of Professor C.U.M. Smith who, in addition to a wide erudition on the history of the neurosciences, cultivated a particular interest on the present frontiers of this discipline, like the subject that runs through the pages of this volume. I am also indebted to Professors Patrizia Calaminici, Bogdan Mielnik, and José Luis Díaz for their critical comments from their different perspectives on preliminary versions of this paper. Residual errors in the text are, of course, the author's sole responsibility. Special thanks are given to Lourdes Ruiz-Zamarripa for her excellent help in retrieving a good deal of bibliography. Recognition is also due to the Center of Consciousness Studies, the Rockefeller University Press, Siglo XXI Editores, and the Universidad Nacional Autónoma de México, for kind permissions to reproduce here previously published figures.

References

Bandyopadhyay A (2011) Direct experimental evidence for quantum states in microtubules and topological invariance. Abstracts: toward a science of consciousness, 2011. Stockholm, p. 131, abs. 191. http://www.consciousness.arizona.edu/documents/FullProgramandAbstractsTSC2011Stockholm.pdf

Bell JS (1987) Speakable and unspeakable in quantum mechanics. Cambridge University Press, Cambridge, UK

Bethe A (1900) Ueber die Neurofibrillen in den Ganglienzellen von Wirbelthieren und ihre Beziehungen zu Golginetzen. Archiv für Mikroskopische Anatomie 55:513–558

Boerhaave H (1742–1746) Dr. Boerhaave's academical lectures on the theory of physic, being a genuine translation of his institutes and explanatory comment, collated and adjusted to each other, as they were dictated to his students at the University of Leyden (Institutiones medicae) (6 vols.). W. Innys, London

Chalmers D (1995) Facing up to the problem of consciousness. J Conscious Stud 2:200–219

Chang C, Hsieh Y-W, Lesch BJ, Bargmann CI, Chuang CF (2011) Microtubule-based localization of a synaptic calcium-signaling complex is required for left-right neuronal symmetry in *C. elegans*. Development 138:3509–3518

Cingolani LA, Goda Y (2008) Actin in action: the interplay between the actin cytoskeleton and synaptic efficacy. Nat Rev Neurosci 9:344–356

Craddock TJA, Tuszynski JA (2010) A critical assessment of the information processing capabilities of neuronal microtubules using coherent excitations. J Biol Phys 36:53–70

Craddock TJA, Beauchemin C, Tuszynski JA (2009) Information processing mechanisms in microtubules at physiological temperature: model predictions for experimental tests. BioSystems 97:28–34

Craddock TJA, Tuszynski JA, Priel A, Freedman H (2010) Microtubule ionic conduction and its implications for higher cognitive functions. J Integr Neurosci 9:103–122

Craddock TJA, Tuszynski JA, Hameroff S (2012) Cytoskeletal signaling: is memory encoded in microtubule lattices by CaMKII phosphorylation? PLoS Comput Biol 8:e1002421. doi:10.1371/journal.pcbi.1002421

Davis PCW (2004) Does quantum mechanics play a non-trivial role in life? BioSystems 78:69–79

Ebner M, Hameroff S (2011) Lateral information processing by spiking neurons: a theoretical model of the neural correlate of consciousness. Comput Intell Neurosci 2011:1–17

Engel GS, Calhoun TR, Read EL, Ahn T-K, Mancal T, Cheng Y-C, Blankenship RE, Fleming GR (2007) Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. Nature 446:782–786

Faber J, Portugal R, Rosa LP (2006) Information processing in brain microtubules. BioSystems 83:1–9

Frampton M (2008) Embodiments of will – anatomical and physiological theories of voluntary animal motion from Greek antiquity to the Latin middle ages, 400 B.C. – A.D. 1300. VDM Verlag Dr. Müller, Chicago/Saabrücken

Freud S (1882) Über den Bau der Nervenfasern und Nervenzellen beim Flusskrebs. Sitzungsberichte Akademie der wissenschaften Wien mathematisch-naturwissenschaftliche Klasse 85:9–46

Frixione E (1983) Firm structural associations of migratory pigment granules with microtubules in crayfish retinula cells. J Cell Biol 96:1258–1265

Frixione E (2000) Recurring views on the structure and function of the cytoskeleton: a 300 year epic (Review). Cell Motil Cytoskeleton 46:73–94

Frixione E (2003) Sigmund Freud's contribution to the history of the neuronal cytoskeleton. J Hist Neurosci 12:11–24

Frixione E (2007) Un Modelo Cuántico de la Conciencia. In: Frixione E (ed) Conciencia – Nuevas Perspectivas en Torno a un Viejo Problema. Universidad Nacional Autónoma de México – Siglo XXI Editores, México, pp 81–89

Frixione E (2009) Cajal's second great battle for the neuron doctrine: the nature and function of neurofibrils. Brain Res Rev 59:393–409

Frixione E, Hernández M (2011) Structural organization of cells – the cytoskeleton. In: Moo-Young M (ed) Comprehensive biotechnology, vol 1, 2nd edn. Elsevier, Amsterdam, pp 367–381

Fröhlich H (1968) Long-range coherence and energy storage in biological systems. Int J Quantum Chem 2:641–649

Galvani L (1791) De viribus electricitatis in motu musculari commentarius [Commentary on the Effect of Electricity on Muscular Motion]. De Bononiensi Scientiarum et Artium Instituto atque Academia Commentarii 7:363–418

Grush R, Churchland PS (1995) Gaps in Penrose's toilings. J Conscious Stud 2:10–29

Hagan S, Hameroff SR, Tuszynski JA (2002) Quantum computation in brain microtubules: decoherence and biological feasibility. Phys Rev E 65:061901

Haikonen POA (2010) Quasi-quantum computing in the brain? Cognit Comput 2:63–67

Hameroff SR (1994) Quantum coherence in microtubules: a neural basis for emergent consciousness? J Conscious Stud 1:91–118

Hameroff SR (1998a) "Funda-mentality": is the conscious mind subtly linked to a basic level of the universe? Trends Cogn Sci 2:119–124

Hameroff S (1998b) Quantum computation in brain microtubules: the Penrose-Hameroff "Orch OR" model of consciousness. Philos Trans R Soc Lond 356:1869–1896

Hameroff S (2001) Consciousness, the brain, and spacetime geometry. Ann N Y Acad Sci 929:74–104

Hameroff SR (2004) A new theory of the origin of cancer: quantum coherent entanglement, centrioles, mitosis, and differentiation. BioSystems 77:119–136

Hameroff SR (2006) The entwined mysteries of anesthesia and consciousness – is there a common underlying mechanism? Anesthesiology 105:400–412

Hameroff SR (2007) The brain is both neurocomputer and quantum computer. Cognit Sci 31:1035–1045

Hameroff S (2010) The "conscious pilot" —dendritic synchrony moves through the brain to mediate consciousness. J Biol Phys 36:71–93

Hameroff S, Penrose R (1996) Conscious events as orchestrated space-time selections. J Conscious Stud 3:36–53

Hameroff SR, Watt RC (1982) Information processing in microtubules. J Theor Biol 98:549–561

Hameroff SR, Smith SA, Watt RC (1986) Automaton model of dynamic organization in microtubules. Ann N Y Acad Sci 466:949–952

Hameroff SR, Nip A, Porter M, Tuszynski J (2002) Conduction pathways in microtubules, biological quantum computation, and consciousness. BioSystems 64:149–168

Hameroff SR, Craddock TJA, Tuszynski JA (2010) "Memory bytes" — molecular match for CaMKII phosphorylation encoding of microtubule lattices. J Integr Neurosci 9:253–267

Hejnowicz Z (1970) Propagated disturbances of transverse potential gradient in intracellular fibrils as the source of motive forces for longitudinal transport in cells. Protoplasma 71:343–364

Hepp K (1998) Toward the demolition of a computational quantum brain. In: Blanchard P, Jadczyk A (eds) Quantum future, vol 517, Lecture notes in physics. Springer, Berlin, pp 92–104

Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 117:500–544
Hollenbeck PJ, Saxton WM (2005) The axonal transport of mitochondria. J Cell Sci 118:5411–5419
Hooft G, Susskind L, Witten E, Fukugita M, Randall L, Smolin L, Stachel J, Rovelli C, Ellis G, Weinberg S, Penrose R (2005) A theory of everything? Nature 433:257–259
Horgan J (1997) The end of science. Facing the limits of knowledge in the twilight of the scientific age. Broadway Books, New York
Jaeken L (2007) A new list of functions of the cytoskeleton. Life 59:127–133
Jibu M, Hagan S, Hameroff SR, Pribram KH, Yasue K (1994) Quantum optical coherence in cytoskeletal microtubules: implications for brain function. BioSystems 32:195–209
Kandel ER, Schwartz JH, Jessell TM (1995) Essentials of neural science and behavior. Appleton and Lange, Stanford
Kendon VM, Nemoto K, Munro WJ (2010) Quantum analogue computing. Philos Trans R Soc Lond 368:3609–3620
Koch C, Hepp K (2006) Quantum mechanics in the brain. Nature 440:611–612
Kurzweil R (2005) The singularity is near – when humans transcend biology. Viking, New York
Lazar JW (2010) Acceptance of the neuron theory by clinical neurologists of the late-nineteenth century. J Hist Neurosci 19:349–364
Libet B, Wright EW Jr, Feinstein B, Pearl DK (1979) Subjective referral of the timing for a conscious sensory experience. Brain 102:193–224
Litt A, Eliasmith C, Kroon FW, Weinstein S, Thagard P (2006) Is the brain a quantum computer? Cognit Sci 20:1–11
Llinás R, Ribary U, Contreras D, Pedroarena C (1998) The neuronal basis of consciousness. Philos Trans R Soc Lond 353:1841–1849
London F, Bauer E (1939) La Théorie de l'Observation en Méchanique Quantique. Hermann, Paris
McKemmish LK, Reimers JR, McKenzie RH, Mark AE, Hush NS (2009) Penrose-Hameroff orchestrated objective-reduction proposal for human consciousness is not biologically feasible. Phys Rev E 80:021912
Merriam EB, Lumbard DC, Viesselmann C, Ballweg J, Stevenson M, Pietila L, Hu X, Dent EW (2011) Dynamic microtubules promote synaptic NMDA receptor-dependent spine enlargement. PLoS One 6:e27688. doi:10.1371/journal.pone.0027688
Mielnik B, Rosas-Ortiz O (2009) Quantum mechanical laws. In: Morán-López JL, Hess OP (eds) Fundamentals of physics, vol 1, Encyclopedia of life support systems (UNESCO). Eolss, Oxford, pp 255–326. Also available as eBook at http://www.eolss.net/outlinecomponents/Fundamentals-Physics.aspx
Moravec H (1990) Mind children. The future of robot and human intelligence. Harvard University Press, Cambridge, MA/London
Penrose R (1989) The emperor's new mind. Concerning computers, minds, and the laws of physics. Oxford University Press, Oxford
Penrose R (1994a) Shadows of the mind. A search for the missing science of consciousness. Oxford University Press, Oxford
Penrose R (1994b) Mechanisms, microtubules and the mind. J Conscious Stud 1:241–249
Penrose R (1996) On gravity's role in quantum state reduction. Gen Relat Gravit 28:581–600
Penrose R (2001) Consciousness, the brain, and spacetime geometry: an addendum. Some new developments on the Orch OR model of consciousness. Ann N Y Acad Sci 929:105–110
Penrose, Hameroff RS (1995) What 'gaps'? Reply to Grush and Churchland. J Conscious Stud 2:98–111
Penrose R, Hameroff S (2011) Consciousness in the universe: neuroscience, quantum space-time geometry and Orch OR theory. http://www.leecharleskelley.com/images/Consciousness_in_the_Universe.pdf
Pizzi R, Strini G, Fiorentini S, Pappalardo V, Pregnolato M (2010) Evidences of new biophysical properties of microtubules. In: Kwon SJ (ed) Artificial neural networks. Nova Science Publishers, New York, ch. 20

Priel A, Ramos AJ, Tuszynski JA, Cantiello HF (2006) A biopolymer transistor: electrical amplification in microtubules. Biophys J 90:4639–4643
Ramón y Cajal S (1899–1904) Textura del sistema nervioso del hombre y de los vertebrados: Estudios sobre el plan estructural y composición histológica de los centros nerviosos, adicionados de consideraciones fisiológicas fundadas en los nuevos descubrimientos (2 vols.). Nicolás Moya, Madrid
Remak R (1843) Ueber den Inhalt der Nervenprimitivröhren. Archiv für Anatomie, Physiologie und wissenschaftliche Medizin, p 197
Remak R (1844) Neurologische Erläuterungen. Arch für Anat Physiol wiss Med 1844:463–472
Rosa LP, Faber J (2004) Quantum models of the mind: are they compatible with environmental decoherence? Phys Rev E 70:031902
Saha AA, Craddock TJA, Tuszynski JA (2012) An investigation of the plausibility of stochastic resonance in tubulin dimmers. BioSystems 107:81–87
Sataric MV, Sekulic D, Zivanov M (2010) Solitonic ionic currents along microtubules. J Comput Theor Nanosci 7:2281–2290
Schrödinger E (1944/1994) What is life? The physical aspect of the living cell. In: Adler MJ (ed) Great books of the western world, vol 56. Encyclopædia Britannica, Chicago, pp 461–504
Schultze M (1870–1871) The general characters of the structures composing the nervous system. In: Stricker S (ed), Power H (transl) Manual of human and comparative histology. Sydenham Society, London
Schwartz JM, Stapp HP, Beauregard M (2005) Quantum physics in neuroscience and psychology: a neurophysical model of mind-brain interaction. Philos Trans R Soc Lond 360:1309–1327
Shepherd GM (1991) Foundations of the neuron doctrine. Oxford University Press, New York/ Oxford
Smith CUM (2006) The ‘hard problem’ and the quantum physicists. Part 1: The first generation. Brain Cogn 61:181–188
Smith CUM (2009) The ‘hard problem’ and the quantum physicists. Part 2: Modern times. Brain Cogn 71:54–63
Smith CUM, Frixione E, Finger S, Clower W (2012) The animal spirit doctrine and the origins of neurophysiology. Oxford University Press, New York
Tegmark M (2000a) The importance of quantum decoherence in brain processes. Phys Rev E 61:4194–4206
Tegmark M (2000b) Why the brain is probably not a quantum computer. Inform Sci 128:155–179
Thagard P (2002) How molecules matter to mental computation. Philos Sci 69:429–446
van der Loos H (1967) The history of the neuron. In: Hydén H (ed) The neuron. Elsevier, Amsterdam, pp 1–47
van Swieten G (1744) The commentaries upon the aphorisms of Dr. Hermann Boerhaave, the late learned professor of Physick in the University of Leyden, concerning the knowledge and cure of the several diseases incident to human bodies, Trans into English, vol I. John and Paul Knapton, London
von Apáthy S (1897) Das leitende Element der Nervensystems und seine topographischen Beziehungen zu den Zellen. Mittheilungen aus der zoologische Station zu Neapel 12:495–748
Ward BK (2014) Is there a link between quantum mechanics and consciousness? In: Smith CUM, Whitaker H (eds) Brain, mind, and consciousness in the history of neuroscience, philosophy and theory of the life sciences 6, Springer, New York, pp 273–302
Wheeler J, Zurek W (1983) Quantum theory and measurement. Princeton University Press, Princeton
Wiener N (1948) Cybernetics or control and communication in the animal and the machine. Hermann, Paris
Wigner EP (1962) Remarks on the mind-body question. In: Good IJ (ed) The scientist speculates: an anthology of partly-baked ideas. London: Heinemann, pp 284–302
Woolf NJ (2006) Acetylcholine, cognition, and consciousness. J Mol Med 30:219–222
Woolf NJ, Hameroff SR (2001) A quantum approach to visual consciousness. Trends Cogn Sci 5:472–478

Chapter 17
Zombie Dawn: Slavery and the Self in the Twenty-First Century

David Hawkes

17.1 Dead Souls

The ideological need to explain consciousness by reference to the material brain long precedes the technical procedures that might make such explanations plausible. The history of phrenology, dedicated to the proposition that, as Hegel cruelly put it, 'the spirit is a bone,' shows how strongly many people in the nineteenth century felt compelled to identify the self with matter at a time when the technology to actually study the brain was far from existing. This need defies not only technique but also logic. Phrenology looks bizarre today, but the compulsions that produced it have only intensified, and it is no more rational to equate an idea with a neurological reaction in the brain, or to speak of a 'gay gene,' or an 'aggression module,' than it is to identify bumps on the skull with criminal tendencies. This essay suggests some reasons behind the evidently imperative desire to find physical causes for mental events.

Perhaps the most disconcerting extrapolation from the ascription of mental events to the brain's physical processes is the challenge it seems to pose to the existence of a unified self, subject or soul which experiences those events. The existence of such an autonomous, non-material subject has long been under question in humanist philosophy and literary criticism as well as in cognitive neuroscience. Friedrich Nietzsche was among the first to criticize the self as a popular illusion:

> A quantum of force is equivalent to a quantum of drive, will, effect—more, it is nothing other than precisely this very driving, willing, effecting, and only owing to the seduction of language (and of the fundamental errors of reason that petrified in it) which conceives and misconceives all effects as conditioned by something that causes effects, by a "subject," can it appear otherwise. For just as the popular mind separates the lightning from its flash and takes the latter for an action, for the operation of a subject called lightning, so popular morality also separates strength from expressions of strength, as if there were a neutral

D. Hawkes (✉)
Department of English, Arizona State University, Phoenix, AZ, USA
e-mail: David.Hawkes@asu.edu

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_17,

> substratum behind the strong man, which was free to express strength or not to do so. But there is no such substratum; there is no “being” behind doing, effecting, becoming; “the doer” is merely a fiction added to the deed—the deed is everything. (110)

Although Nietzsche’s anti-humanism was loudly extolled by Fascists, for three or four decades after the Second World War, this materialist, anti-essentialist approach to subjectivity became associated with the political Left. The unified subject was frequently denounced as a tool of patriarchal oppression. The original forger of this connection was Louis Althusser, who identified the essentialist concept of the subject as simultaneously the source and the consequence of all ideological error:

> I say: the category of the subject is constitutive of all ideology, but at the same time and immediately I add that the category of the subject is only constitutive of all ideology insofar as all ideology has the function (which defines it) of ‘constituting’ concrete individuals as subjects.[1]

According to Althusser, whose ideas were widely disseminated by his pupil Michel Foucault in the 1970s, the autonomous, unified self was not merely an ideological illusion; it was the *archetype* of ideological illusion, and the source of all other ideological illusion. It is because we experience ourselves as substantive essences that we perceive all kinds of similarly illusory essences lurking beneath empirical appearances. For Foucault and the legions inspired by him, the Myth of the Self revealed a pernicious tendency to perceive essences lurking beneath appearances, to ascribe an ulterior cause and significance to apparent phenomena. By the end of the twentieth century, such arguments converged with a tendency among neuroscientists to deny the existence of any unified subject on empirical grounds. The sudden appearance, dramatic spread, and widespread appeal of such a historically anomalous and counter-intuitive theory as the non-existence of the human subject surely demands explanation.

There are some striking convergences between the humanities and the sciences on this issue. Richard Dawkins’s theory of ‘memes’ closely resembles such post-structuralist concepts as Roland Barthes’s ‘mythemes,’ and the classic texts of eliminative materialism, like Daniel Dennett’s *The Intentional Stance* (1989) echo literary controversies regarding authorial intention:

> Treating a complex, moving entity as a single-minded agent is a magnificent way of seeing pattern in all the activity; the tactic comes naturally to us, and is probably even genetically favored as a way of perceiving and thinking. But when we aspire to a science of the mind, we must learn to restrain and redirect those habits of thought, breaking the single-minded agent down into miniagents and microagents (with no single Boss). (458)

The logical and rhetorical connections with such post-structuralist thinkers as Foucault and Derrida are striking. In ‘What is an Author?’ (1969) Foucault writes:

> Critics doubtless try to give this intelligible being a realistic status, by discerning, in the individual, a ‘deep’ motive, a ‘creative’ power, or a ‘design,’ the milieu in which writing

[1] Althusser 1971.

> originates. Nevertheless, these aspects of an individual which we designate as making him an author are only a projection, in more or less psychologizing terms, or the operations that we force texts to undergo, the connections that we make, the traits that we establish as pertinents, the continuities that we recognize, or the exclusions that we practice.[2]

Post-structuralists are linguistic determinists, who regard the subject as the product of language, while cognitive neuroscientists are materialist determinists, who regard the subject as the result of the brain's physical operations, but the two critiques of subjectivity are entirely compatible. In *Kinds of Minds,* Dennett (1997) declares that in the absence of physical evidence for the existence of a unifying principle, a 'Cartesian puppeteer,' it must be treated as mythical. A person's identity 'just *is* this organization of all the competitive activity between a host of competences that your body has developed' (155–156). It is something of a mystery that people should continue to 'ask for more,' to hanker after a 'central knower' which might exceed the sum of its material parts.

For Dennett and similar eliminative materialists, there is no element of a human being prior to its material manifestation, just as for post-structuralist semioticians there is no thought prior to its expression in language. There is thus no point in seeking a text's significance in the mind of its author. Roland Barthes claims that subjectivity is irreducibly textual: 'Did [the author] wish to express himself, he ought at least to know that the inner "thing" he thinks to "translate" is itself only a ready-formed dictionary, its words only explainable through other words, and so on indefinitely.' Writing, he claims, is no longer to be regarded as 'representation,' either of external reality or of the thoughts within the author's mind, but as a 'performative' discourse.

Barthes (1978) refers here to the work of J.L. Austin, who drew a distinction between 'constative' statements, which make truth-claims about the world (such as 'the car is red'), and 'performative' statements, which constitute objective effects on the world (such as 'I now declare you man and wife.') Such statements attain an autonomous power, their effect is independent of the thoughts within the conscious mind that produced them. Austin begins by differentiating between 'constative' statements like 'the man is tall,' and 'performative' statements such as 'I now pronounce you man and wife.' The latter statement is an action rather than a description; it has no ulterior meaning other than the act it carries out. Furthermore, the source of its efficacy lies in the words themselves, not in the conscious intention of the speaker. The couple will objectively be married after the words are pronounced, even if the priest subjectively intends them to remain single.

At first, Austin presents the 'performative' as an exceptional usage, and the 'constative' as the normative instance, but as his analysis develops, he is led inexorably to the logical conclusion that *any* statement constitutes an action: 'When we issue any utterance whatsoever, are we not "doing something?"'[3] The category of the performative turns out to be primary and fundamental, and the category of the constative

[2] Foucault 1969.

[3] Austin 1962.

is ‘parasitical’ upon it. It provides an explanation for how linguistic meaning might operate in the absence of a directing intelligence. As we shall see, however, performative representation has historically been the object of extremely cogent and potent ethical criticism for, among other things, its attack on the concept of ‘truth.’ Performative statements are not either true or false, but successful or unsuccessful. A world in which all statements are performative is the moral universe of pragmatism, where truth is what works, ideas do not exist, appearance is reality, and human beings are identical with their bodies.

There is something intuitively difficult to accept about such a world, as post-structuralists and neuroscientists alike are usually willing to admit. David Chalmers has stated the problem succinctly. Chalmers’s famous ‘zombie’ thought experiment is an application of dialectical logic to the question of consciousness. In the act of identifying any entity as existing, we automatically bring something other than that entity into conceptual existence. Where there is an object, there must be a subject. The act of defining a sphere of the ‘material’ not only proves but actually creates the existence of a non-material sphere. Chalmers approaches the question of consciousness using this logic. If we can conceive of a being which looks and behaves exactly like a human without having any subjective experience—a ‘philosophical zombie’—then by that very act of conception we will have both proved and created the existence of a being that looks and behaves exactly like a human and *does* have subjective experience. We will have proved that physical factors alone cannot explain consciousness:

> Any story about physical process applies equally to me and to my zombie twin. It follows that nothing in that story says why, in my case, consciousness arises…. The very fact that it is logically possible that the physical facts could be the same while the facts about consciousness are different shows us that… there is an explanatory gap between the physical level and conscious experience. (107)

The true significance of the ‘hard problem,’ however, is that it should have arisen in the first place. Chalmers frames the issue from a neuroscientist’s perspective: ‘Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should, and yet it does.’ (203) But to whom does it seem ‘objectively unreasonable?’ Before the eighteenth century, most people would have agreed that any attempt to account for ideas solely by reference to material processes within the brain was manifestly absurd. The notion that our subjective experience transcends our physical experience seemed viscerally obvious and logically incontrovertible. The phenomenon we might call ‘the death of the soul’ is specific to Western modernity. If the idea that we are identical with our bodies has come to seem plausible today, our first problem is to account for that plausibility.

In other words, the *really* hard problem is why neuroscientists find it ‘objectively unreasonable’ that people should have conscious experience. That is a question that neuroscientists themselves are forbidden even to ask. The first principles of their discipline prohibit the location of their thought in its historical or social context. But the intellectual developments that have called the notion of consciousness into question are themselves part of history, and can only be explained with reference to the historical circumstances that have produced them. It is only by locating the materialist approach to consciousness within history that we can begin to understand why it has arisen. This essay offers a tentative step towards such an account.

17.2 Dead Men Working

The history of Western thought shows that the 'philosophical zombie' is no recent invention. Indeed it is an ancient phenomenon, and it has been used throughout history for the purpose which Chalmers ascribes to it. People have always confirmed the existence of their own consciousness, and analyzed its nature, by comparing it with something that, although similar to the possessor of consciousness in all other relevant respects, lacks consciousness. The individual mind, for example, is defined as conscious by comparison with the individual body. Although they are parts of the same individual person, the mind is conscious while the body is not. The mind is conceived as conscious in contrast to the nature of the body. The unconsciousness of the body reveals and describes the consciousness of the mind.

Because it belongs to an individual, however, the human body cannot fulfill the function of Chalmers's zombie at the level of society. It might suffice to convince a single individual of his own conscious experience, but it could assure neither the individual nor others that their experience of selfhood was universal. Only the idea of an entire humanoid being which was identical to actual humans apart from the lack of conscious subjectivity could demonstrate that human beings do possess conscious subjectivity. This is the theoretical function of Chalmers's philosophical zombie in the modern world. Historically however, that function has been performed by slavery.

That is not to say that people believed that individual slaves necessarily lacked conscious experience or subjectivity. That belief was uncommon, at least until the burgeoning of the Atlantic trade infiltrated a racist component into the ideology of slavery. But the condition of slavery *per se* has universally been defined as involving the lack of conscious experience and subjectivity, whether or not that condition applies to the individual slave. Although an enslaved person might possess consciousness *qua* person, *qua* slave he does not. The condition of slavery, as opposed to the empirical nature of the enslaved, has played the role of 'philosophical zombie' throughout human history. Chalmers's version only became necessary when the condition of slavery ceased to exist in empirical reality.

With its unrivaled gift for hypocrisy, our society regards slavery as an egregious ethical transgression. Yet the ancient world possessed a perfectly coherent, and extremely long-lived ethical *rationale* for the institution. It was permissable to buy and sell slaves, to commodify them and treat them as things because, socially speaking, they were already dead. The conventions of warfare were remorseless in their logic. In war, people attempt to kill one another. To surrender was therefore to lose one's life, not by physical extinction, but by giving it up to the power of another.

As the Roman jurist Ulpian noted: 'In every branch of the law, a person who fails to return from enemy hands is regarded as having died at the moment he was captured.'[4] The same logic applied to the other means by which slavery might be imposed—as a substitute for capital punishment, for instance, or to prevent an infant's death by exposure. Slavery was what happened to a person when death was the only alternative and as such it was itself a form of death. Slaves could be treated as things because, socially though not biologically, that is exactly what they were.[5]

Moralistic objections to slavery could therefore be answered by pointing out that the master was actually doing the slave a favor by keeping him biologically alive. Escape through suicide was always a possible option for the slave, and by exercising it he would merely be attaining the state to which fate had already assigned him. By choosing to remain alive once enslaved, the slave was enslaving himself. Furthermore, only a person naturally suited to slavery would make such a choice. Anyone who was by nature a free person could never be enslaved, for they would die first. The ex-slave Epictetus (1916) understood his former condition with a clarity born of experience:

> … those animals only are free which cannot endure capture, but, as soon as they are caught, escape from captivity by death. So Diogenes says that there is one way to freedom, and that is to die content: and he writes to the Persian king, "You cannot enslave the Athenian state any more than you can enslave fishes." "How is that? cannot I catch them?" "If you catch them," says Diogenes, "they will immediately leave you, as fishes do; for if you catch a fish, it dies; and if these men that are caught shall die, of what use to you is the preparation for war?" (4.1)

By electing to remain alive in captivity, a person revealed themselves as a natural slave, and the act of legally enslaving them could therefore be regarded as in accordance with natural justice. Those who could be slaves *should* be slaves. Indeed the capacity for slavery was itself a form of slavery, whether or not it attained legal status. Hegel made the same point more than twenty centuries later:

> If a man is a slave, his own will is responsible for his slavery, just as it is its will which is responsible if a people is subjugated. Hence the wrong of slavery lies at the door not simply of enslavers or conquerors, but of the slaves and the conquered themselves.[6]

What sets slavery apart from other forms of unfree labor is the reduction of the entire person to a commodity. Individual people are qualitatively distinct by nature,

---

[4] Ulpian, *Sabinus* book 35, cit. Watson, 404. Orlando Patterson cites Ali Abd Elwahed: 'all the situations which created slavery were those which commonly would have resulted, either from natural or social laws, in the death of the individual.' Patterson 1982.

[5] See Bradley 1994: 'The victor in battle had the right to kill the vanquished. If, however, the victor spared the vanquished and enslaved him instead, the latter continued to live, but only in a condition of suspended death, at the discretion of the former. The slave's very identity, in fact, now depended on his owner. This was the source of the slave-owner's power.' (26) See also Patterson 1982: 'Perhaps the most distinctive attribute of the slave's powerlessness was that it always originated (or was conceived as having originated) as a substitute for death, usually violent death.' (5)

[6] Hegel 1967 Cit. Buck-Morss 2000, 849.

but they can be made quantitatively equivalent for the purposes of exchange, and this transformation of a person into a commodity is known as 'reification.' The distinction between 'objectification' and 'reification' is subtle but vital. The former implies the reduction of a human being to matter. 'Reification' involves the reduction of a person to a thing, in legal though not necessarily in ontological terms. As a commodity, a slave could be biologically human without being legally so. Commodities may or may not be material, but they are always things.

Orlando Patterson makes the point succinctly in *Slavery and Social Death*: 'the most common conception of the slave among the Romans became, by the end of the republic, that of a thing. The slave was above all a *res, the only human res.*'[7] Patterson offers a detailed analysis of Roman law as it applied to slavery, which was based on the distinction between *persona* and *res.* Because his owner enjoyed *dominium* over him, the slave fell legally into the category of *res.* Reification (transformation into a thing) is thus quite literally the essence of slavery. The concept of *dominium* involves more than simple ownership; it describes the relation between the owner and the owned. As Patterson explains, *dominium* cannot be exercised over an inanimate object: 'It is difficult to explain why the Romans would want to invent the idea of a relation between a person and a thing… unless we understand that, for most purposes, the "thing" on their minds was a slave.' (31) To understand the ancient view of slavery we must distinguish between a material 'object' and a 'thing' in the sense of *res.*

It was this distinction that made the slave the 'philosophical zombie' of antiquity. Although they were clearly human in the biological sense, slaves *qua* slaves did not possess conscious subjectivity. The servile mind (as opposed to the minds of slaves) is consistently described as carnal or, in modern terms, materialist. Jennifer Glancy notes that the 'equation between slaves and bodies actually begins with the lexicon of slavery. The Greek word for body, *to soma,* serves as a euphemism for the person of a slave.' (9) Glancy calls attention to the longevity of this conception, which remained constant throughout classical history. In Rome as in Greece:

> … slavery was identified with the body… Slaveholders in the first century characterized their slaves as bodies, and their treatment of their slaves was commensurate with that characterization. This was equally the case in the fourth century, when Constantine came to power, and a century after that.[8]

[7] Patterson 1982, 32. Almost a century earlier, William Buckland observed that in Roman law 'a slave was a *Res,* and, for the classical lawyers, the only human *Res*.... From the fact that a slave is a *Res* it is inferred, apparently as a necessary deduction, that he cannot be a person. Indeed the Roman slave did not possess the attributes which modern analysis regards as essential to personality.' William Warwick Buckland, *The Roman Law of Slavery* (orig 1908, reprinted CUP, 2010), 3. See also Bradley 2000. Bradley notes that 'the common Greek term for slave, *andeapodon,* "man-footed creature," was built on the foundation of a common term for cattle, namely, *tetrapodon,* "four-footed creature."' (110)

[8] Glancy 2002, 9.

Henri Wallon describes the ancient conception of a slave as 'a body with natural movements but without its own reason, an existence entirely absorbed in another. The proprietor of this thing, the mover of this instrument, the soul and reason of this body, the source of this life, was the master.'[9] Glancy cites Artemidorus's *Oneirocritica:* '"The very man who dreamt that he saw his household slave sick with a fever became ill himself, as one might expect. For the household slave has the same relationship to the dreamer as the body has to the soul."' (9–10)

Such texts gave popular expression to philosophical profundities. According to Plato's internal ethical hierarchy, reason should rule in the soul. The proper function of the passions and the appetites is to serve the ends of reason. But when these lower elements rule over their natural masters the result is psychological tyranny, and a man in whom the slavish elements rule the rational will also be inclined towards political tyranny. This connection between psychological and political servility informed Western culture for millennia. Two thousand years after Plato, John Milton (1959) blamed the English people's preference for monarchy over republicanism on their servitude to their own desires: 'being slaves within doors, no wonder that they strive so much to have the public State conformably govern'd to the inward vitious rule, by which they govern themselves.' Until the American Civil War, pro-slavery advocates like James Shannon utilized the same argument against abolitonists: 'the slaves of ignorance and vice are incapable of self-government.'[10]

Being entirely physical creatures, slaves *qua* slaves lack the power of self-mastery. As William Channing observed, 'power over himself is the last virtue we should expect in the slave.' (69) Being unable to liberate themselves from internal slavery to their physical impulses, slaves had no autonomous subjectivity that could be acknowledged legally. Insofar as the nature of slaves was a matter of legal notice, they were defined in purely economic terms. As Junius P. Rodriguez (1997) explains, in ancient law: 'a slave's level of production was deemed to be far more important than his or her actual person or identity, and accordingly the slave was essentially an economic creature.'[11] The slave is the original *homo economicus,* and the modern incarnation of that figure bears the visible brand of its original. Plato's (1892) *Republic* points out that, since the desires of the body require only money for their satisfaction, this slavish element of the soul is finds its paradigmatic expression in avarice. This lowest element:

> … having many forms, has no special name, but is denoted by the general term appetitive, from the extraordinary strength and vehemence of the desires of eating and drinking and the other sensual appetites which are the main elements of it; also money-loving, because such desires are generally satisfied by the help of money…. If we were to say that the loves and pleasures of this third part were concerned with gain, we should then be able to fall back on a single notion; and might truly and intelligibly describe this part of the soul as loving gain or money.

[9] Cit. Patterson 1982, 4.

[10] Shannon 1849.

[11] Junius P. Rodrigues, *The Encyclopedia of Slavery,* xiii.

A soul that has suffered such an internal slave rebellion will find itself correspondingly demoted in external status. Properly speaking, the body exists to serve the soul. But a person whose appetites serve their reason elevates the concerns of their body over those of their soul. Such a person denies the nature of the soul and thus, insofar as possible, obviates the soul's very existence. As Malcolm Bull remarks: 'If slavery is the rule of the slave's body by a soul of another, then the very possibility of slavery depends upon the slave's body not being governed by the slave's own soul.'[12]

Aristotle describes slaves as are purely sensual beings, 'wholly lacking the deliberative element' (*Politics*, 1260al, 12). Like Plato, he compares the master/slave relation to that of the mind to the body 'that which can foresee with the mind is the naturally ruling and naturally mastering element, while that which can do these things with the body is the naturally ruled and slave' (*Politics*, 1252bl). But Aristotle also raises what seems to us an obvious ethical objection to slavery. Some people, he reports, claim: 'that the rule of a master over slaves is contrary to nature, and that the distinction between slave and freeman exists by law only, and not by nature; and being an interference with nature is therefore unjust.' (cit. Meltzer 1993)

Aristotle does not give the refutation we might expect from a modern liberal, he does not argue that every human being has the right to be free. His moral objection to slavery in practice is that it does not always enslave the right people. The struggle against servility takes place constantly within each of us; only those who win it deserve to be externally free. The injustice lies not in slavery *per se,* but in the imposition of slavery on those who are naturally free and, by inevitable extension, in the liberty of those who are naturally slaves. Aristotle assumes that the naturally free man will devote himself to the proper end of humanity, which is the cultivation of the soul. Such a man cannot be a slave, for a slave must serve his master's ends rather than his own: he must play the role of body to his master's soul. He therefore cannot possess autonomous subjectivity and becomes, rather, a 'property' of somebody else: 'the slave is a part of the master, a living but separated part of his bodily frame.' (19) As we should use our bodies to serve our souls, so the *telos* of the slave is not the maintenance of his own life, but that of his master, to which end he is merely a means:

> The master is only the master of the slave; he does not belong to him, whereas the slave is not only the slave of his master, but wholly belongs to him. Hence we see what is the nature and office of a slave; he who is by nature not his own but another's man, is by nature a slave; and he may be said to be another's man who, being a human being, is also a possession. And a possession may be defined as an instrument of action, separable from the possessor.

The salient characteristic of Aristotle's natural slave is that he is an 'instrument,' a means to someone else's ends. The slave's activity is alienable, 'separable from the possessor,' and does not belong to the person who carries it out. This condition of alienation dehumanized the slave, for it separated his life from his self. The life of a slave is not his own.

[12] Malcolm Bull 2000, *Seeing things Hidden: Apocalypse, Vision and Totality,* 210.

17.3 Dead on Arrival

As anyone who has ever worked for a wage knows very well, this is true of proletarians[13] as well as of slaves. A wage-worker's time is not his own while he is at work. During work hours, the worker serves his employer's ends. The proletarian sells himself piecemeal, while the slave is sold outright, but by the logic of the classical world both are subjected to the process of commodification, with all of its attendant psychological effects.[14] In Cicero's (1991) *De Officiis,* proletarians are degraded to the level of slaves on the grounds that they sell their time, rather than the products of their labor: 'Illiberal too, and mean, are the employments of all who work for wages, whom we pay for their labor and not for their art; for in their case their very wages are the warrant of their slavery.' (1:150–151)[15] David Graeber finds that:

> … wage labor contracts appear to have developed from within the institution of slavery in many times and places, from ancient Greece to the Malay and Swahili mercantile city states of the Indian Ocean. Historically, I think one can say wage labor, at least considered as a contractual arrangement, emerged from slavery….[16]

According to Graeber, this historical connection reflects a fundamental conceptual affinity. A proletarian, like a slave, is under the control of another's will:

> … what one buys when one buys a slave is the sheer capacity to work, which is also what an employer acquires when he hires a laborer. It's of course this relation of command that causes free people in most societies to see wage labor as analogous to slavery, and hence to try as much as possible to avoid it. (79)

The concept of freedom itself develops as the dialectical antithesis of slavery. Graeber points out that '[t]he modern ideal of political liberty… has historically

---

[13] A 'proletarian' is anyone who works for a wage; the term does not necessarily imply poverty.

[14] See Finley 1973: 'The very idea of wage labor…. requires the abstraction of a man's labor from both his person and the product of his work. When one purchases a product from an independent craftsman, whether he is free or a slave with a *peculium,* one has not bought his labor but the object, which he had produced in his own time and under his own conditions of work. But when one hires labor, one purchases an abstraction, labor-power, which the purchaser then uses at a time and under conditions which he, the purchaser, not the "owner" of the labor-power determines (and for which he normally pays after he has consumed it.)' (65)

[15] David Graeber 2001 points out that in ancient China, slavery and wage labor were: 'two phenomena that, as so often in the ancient world, largely overlapped: the common phrase for workers used in texts from the period was (dasa—karmakara), 'slave—hireling' with the assumption that slaves and laborers worked together and were barely distinguishable.' *Towards an Anthropological Theory of Value*, (Palgrave 2001), 428n38.

[16] Graeber 2007, 5. Graeber also cites Friedman's conclusion that 'slavery in Classical Greece is a complex affair involving wage, interest and profit in an elaborate market system that appears to have had cyclical properties of expansion and contraction. This was, in other words, a form of capitalism that is not so different from the more obvious varieties in the modern world.' (68)

tended to emerge from societies with extreme forms of chattel slavery (Pericles' Athens, Jefferson's Virginia), essentially as a point of contrast.' (79) When slavery is replaced by wage labor, however, the contradiction between bondage and liberty is internalized:

> We are dealing with the same terms, differently arranged, so that rather than one class of people being able to imagine themselves as absolutely 'free' because others are absolutely unfree, we have the same individuals moving back and forth between these two positions over the course of the week and working day. (80)

According to Graber's logic, the relation between bondage and liberty is internalized in the modern world. Society is no longer divided into the enslaved and the free; in a system of wage labor, everyone is both free and enslaved. In the ancient world, wage-labor was a small part of the economy. Since the eighteenth century, however, wage labor has spread rapidly, and in 'developed' societies it is virtually universal today. We are all proletarians now, but we are yet to come to ideological terms with this unprecedentedly sudden transformation in human life.[17] That is why we are tempted to mistake the peculiar psychological consequences of proletarianization for immutable truths about the self.

As we have seen, the first rationalizations of slavery treated it as an alternative to death in battle. Understood in this way, slavery could seem just even to its victims. In Aphra Behn's *Oroonoko, the Royal Slave,* the eponymous hero declares that slavery would be endurable if it resulted from military defeat:

> Have they vanquished us nobly in fight? Have they won us in honourable battle? And are we by the chance of war become their slaves? This would not anger a noble heart; this would not animate a soldier's soul: no, but we are bought and sold like apes or monkeys, to be the sport of women, fools and cowards. … (62)

In the mid-seventeenth century, Behn's hero still adheres to classical concepts: the insult does not lie in slavery *per se* but in commodification. As a warrior, Oroonoko understands the justice of force. As a paradoxically 'royal' slave, however, he cannot submit to the reification imposed upon a commodity. For Behn's (2003) contemporary Thomas Hobbes, on the other hand, it was precisely the slave's ability to conceive of himself as an alienable commodity that justified his enslavement:

> It is not therefore the victory, that giveth the right of dominion over the vanquished, but his own covenant. Nor is he obliged because he is conquered; that is to say, beaten, and taken, or put to flight; but because he cometh in, and submitteth to the victor. (405)

It is not the fact of defeat that justifies slavery, but the agreement of the defeated to be enslaved, even if that consists only in his refusal of suicide. If slavery was a

[17] As Marx 1993 put it in the *Grundrisse* (1865): 'The direct transition from the African's fetish to Voltaire's supreme being, or from the hunting gear of a North American savage to the capital of the Bank of England, is not so absurdly contrary to history as is the transition from Bastiat's fisherman to the wage labourer.' (chapter 17, cclxxvi)

contractual relation it might be voluntary, and this provides Hobbes with his rationalization of wage labor.[18] It is notable that Hobbes (2003) does not imagine that wage labor could be justified in the absence of a justification of slavery, for like every commentator since Aristotle, he still conceives of the former as a sub-species of the latter.

The difference between Hobbes and classical theorists of slavery like Aristotle is that the thinkers of the ancient world found it impossible to conceive of a fully human being who was also a commodity. A person might be one or the other but not both. This was because the process of commodification robs anything of its nature, and imposes a merely conventional, artificial significance upon it. A commodity is not just what it is; it is also what it is worth, what it represents. This 'exchange-value' does not arise out of the commodity's inherent properties, it is imposed from outside. To the degree that anything is regarded as a commodity, its nature is obliterated. To the degree that a human being is commodified, therefore, his human nature is occluded. He becomes a simulacrum of a human being, who may appear to behave in a reasonable manner, but who lacks the rational, independent consciousness that defines a free man.

In *The Wealth of Nations*, Adam Smith observes that money does not automatically bestow political or military power on its possessor. What it does automatically bring him (what it actually is) is the power to direct the activity of other people:

> The power which that possession immediately and directly conveys to him, is the power of purchasing, a certain command over all the labour, or over all the produce of labour which is then in the market. His fortune is greater or less, precisely in proportion to the extent of, his power; or to the quantity either of other men's labour, or, what is the same thing, of the produce of other men's labour, which it enables him to purchase or command. The exchangeable value of everything must always be precisely equal to the extent of this power which it conveys to its owner. (37)

Financial value represents alienated labor-power that has been separated from the person who performs it by being translated into the symbolic form of money. Money is nothing but human labor-power—which is to say human time, human life—in objective form. It is the reification of the human subject. Smith (1976) saw the significance of this point clearly enough to repeat it on several occasions: 'The value of any commodity to the person who possesses it [and wishes to exchange it] is equal to the quantity of labour which it enables him to purchase or command. Labour, therefore, is the real measure of the exchangeable value of all commodities.' (47)

How did this fairly obvious truth remain unremarked by the greatest thinkers of the ancient world? Aristotle certainly understood that the exchange of qualitatively

---

[18] In the words of Tommy Lee Lott: 'The overlap of slavery and servitude… allows the natural rights theories of Hobbes and Locke to condone benign forms of slavery. Hence, such theories cannot provide an adequate ground on which to condemn slavery as inherently evil.' Introduction to Tommy Lee Lott (ed.), *Subjugation and Bondage: Critical Essays on Slavery and Social Philosophy,* (Routledge, 1998), xvii.

different goods involves the imposition on their natural 'use-value' of an artificial equivalent, a common denominator: an 'exchange-value.' He was unable to see, however, that this 'exchange-value' must represent human labor-power.[19] This blind spot was due to the fact that he did not conceive of human labor-power as valuable—not because it was worthless, but because it was not the sort of thing that could be assigned a value. He did not imagine labor-power as a separate commodity to be traded in isolation from the rest of the person who performs it. Only as wage-labor replaced slavery did it become clear that labor itself had (or rather was) value. Only as it became clear that exchange-value, in the symbolic manifestation of money, represented alienated labor-power did it become possible to understand that capital, the power that rules the world, is nothing more than human activity in externalized, objective form.

The vast majority of people in capitalist societies live by selling their time, their lives, their selves, on a daily basis: they are proletarians. According to Aristotle's canonical description, proletarians are piecemeal slaves. It is true that they sell their labor by the hour, not by the lifetime, and they may even experience this as a voluntary bargain. Yet Aristotle defines slavery as human activity directed towards an end that is not proper to the actor but to another. The slave's actions are not carried out for the purposes of the slave, but for those of the master, and they are therefore not the actions of a conscious human subject. By that definition, which was universally accepted for thousands of years, a proletarian is a slave.

Since almost everyone in today's Western and Westernized societies is a proletarian in the crucial, definitive sense that they exchange a portion of their time for money every day, almost everyone in contemporary society is a slave. It is therefore not surprising to find that almost everyone thinks like a slave. The predominant characteristics of the 'postmodern condition,' as described by 'post-humanist' philosophers and cognitive neuroscientists alike, meet the classical definitions of slavery. The central element in such definitions is the destruction, or simply the absence, of any autonomous, spiritual or non-material self.

As we have seen, however, a natural slave is quite distinct from a legal slave. Slavery certainly produces a servile subject-position, but it is not necessarily occupied by a servile subject.[20] In *The Souls of Black Folks,* W.E.B. Dubois provides a compelling analysis of the effect of slavery on subjectivity, arguing that commodification introduces a breach between the outer and the inner self, a systematic disjunction between appearance and essence. The commodification of anything introduces a split between its natural qualities and its artificial exchange-value. In the case of a human commodity, whether the life of a slave or the

[19] See Marx 1967.

[20] Although it certainly can be, as Walter Johnson comments: 'To the social death experienced by those torn from their histories and identities and the physical death they faced in the killing fields of the lower South must be added the psychic deaths—the "soul murder"—that left many of the trade's victims with little will to resist.' Johnson 1999.

40 h per week of a 'free' proletarian, an external, objective identity is superimposed upon the subject. Walter Johnson describes the destructive effect of the imposition of exchange-value on the slave's identity: 'though they were seldom priced, slaves' values always hung over their heads… any slave's identity could be disrupted as easily as a price could be set…'[21]

The theory of natural slavery assumes that, as a general rule, slavery will be beneficial to the slave as well as to the master. Logically speaking, however, the slave is the master's dialectical antithesis: each category defines itself against the other and thus brings the other into being. The great class struggles of the nineteenth and early-twentieth centuries were ignited when labor-power, incarnated in the 'proletariat,' recognized that capital, embodied in the 'bourgeoisie,' was its own objectified form. Yet the practical demands of 'actually existing socialism' in both Communist states and the Western labor movement generally conspired to ignore the identification of slavery with wage-labor. Marx's son-in-law Paul Lafargue was one of very few thinkers to notice its political implications. In *The Right to be Lazy,* Lafargue (1883) recalls that:

> The Greeks in their era of greatness had only contempt for work: their slaves alone were permitted to labor: the free man knew only exercises for the body and mind… The philosophers of antiquity taught contempt for work, that degradation of the free man, the poets sang of idleness, that gift from the Gods…. Proletarians, brutalized by the dogma of work, listen to the voice of these philosophers, which has been concealed from you with jealous care: A citizen who gives his labor for money degrades himself to the rank of slaves. (12)

The final sentence is from Cicero (1991), who understood the proletarian's sale of time as indistinguishable from the commodification of the slave. Nor is the connection merely theoretical. The conditions of proletarian life—which is to say the conditions of almost everybody's working life in the modern Western world—are directly descended from the condition of slavery. This fact lurks unquietly in the attic of formalized labor relations. C.L.R. James was the first to observe that rationalized techniques for the direction of human beings towards ends that they do not naturally wish to pursue were first developed on slave plantations. Bill Cooke has recently pointed out that antebelleum slavery was 'managed according to classical management and Taylorian principles,' so that the subject of slavery is 'of intrinsic, but hitherto denied, relevance to [modern] management studies.'[22] This throws an unappealing light on the concept of 'management' itself. Robbins defines it as 'the process of getting activities completed efficiently with and through other people'

[21] Johnson 1999, 19.

[22] Cooke 2003. Cooke cites Chandler's observation that as the first salaried manager in the USA, 'the plantation overseer was an important person in American economic history. The size of this group (in 1850 overseers numbered 18,859) indicates that many planters did feel that they needed full time assistance to carry out their managerial tasks' (1977, p. 64) He also cites Smith's claim that '"the way slaveholders organized their workforce, the way they treated their bondpeople, their heavy involvement in the market economy, and their drive for profit made them much more capitalist" than historians have previously acknowledged.'

(1993), which is to say that 'management' is the art of treating people as instruments, directing their activities towards ends chosen by others.[23] Cooke's conclusion is unequivocal:

> The imprint of slavery in contemporary management can be seen in the ongoing dominance from that time of the very idea of the manager with a right to manage. It can also be seen in the specific management ideas and practices now known as classical management and scientific management which were collated and re-presented with these labels within living memory of the abolition of US slavery.

It is not only physical activity that is reified by wage labor; psychological processes are subjected to the same effect. Early capitalism found ways to express subjective experiences like 'credit,' 'confidence' or 'goodwill' in figurative, financial form, and to allow these representations to function autonomously, acquiring and losing market value as commodities. In recent years, this process has expanded into more or less brazen attempts to remold the entire human personality in accordance with the demands of capital. The philosophy behind most such endeavors can be traced to Frederick Winslow Taylor's *The Principles of Scientific Management*. The techniques by which Taylor sought to rationalize production are too well-known to bear repeating here. They have been honed and refined throughout the twentieth century, most notably in the 'Lean Manufacturing' philosophy developed at Toyota's car plants during the 1970s. Their aim is to eliminate any aspect of the individual worker's subjective behavior (which Taylor calls 'initiative'), that conflicts with the imperatives of profit. Thus does money, the objectified form of human activity, determine the subjective form of human activity. In the words of Georg Lukacs:

> If we follow the path taken by labour in its development from the handicrafts via cooperation and manufacture to machine industry we can see a continuous trend towards greater rationalisation, the progressive elimination of the qualitative, human and individual attributes of the worker…. With the modern 'psychological' analysis of the work-process (in Taylorism) this rational mechanisation extends right into the worker's 'soul': even his psychological attributes are separated from his total personality and placed in opposition to it so as to facilitate their integration into specialised rational systems and their reduction to statistically viable concepts. (88)

In recent decades, Taylor's methodology has been applied beyond the sphere of production, into the processes of exchange and consumption. Taylor himself aspired

---

[23] It is never the demanding nature of the slave's work that is considered degrading, but the fact that it is not directed towards ends chosen by himself. The most brutal kinds of drudgery can be regarded as ennobling if they are undertaken for freely-chosen ends. As Shakespeare's Prince Ferdinand puts it in *The Tempest:*

> … some kinds of baseness
> Are nobly undergone and most poor matters
> Point to rich ends. This my mean task
> Would be as heavy to me as odious, but
> The mistress which I serve quickens what's dead
> And makes my labours pleasures (3.1.1-7)

to extend scientific management to 'the work of our salesmen,' and attempts by his followers to implement such techniques in the sphere of exchange became known as 'sales engineering.' In its most recent, postmodern form, Taylorism is even applied to what were once regarded as non-economic aspects of life. We have become used to being enjoined to 'manage' every aspect of our behavior, and indeed of our thoughts, in a fashion directly inspired by Taylor's prescriptions for the workplace. We are now being encouraged to replace our subjective identities with 'brands' inspired by the demands of the marketplace. This colonization of 'leisure' by 'work' represents a more profound penetration of capital into the psyche than even Taylor could have envisaged.

17.4 Day of the Dead

During the nineteenth century, as the challenges to slavery grew irresistible, several thinkers warned that the psychological struggle against reification was no less significant. We have already seen how, in *The Souls of Black Folks,* W.E.B. DuBois describes the effect of enslavement on subjectivity in terms reminiscent of Aristotle and Plato. The most deleterious of those effects is the destruction of the unified subjectivity. The slave leads a 'double life, with double thoughts, double duties, and double social classes', and, in consequence, 'ever feels this twoness,—an American, a Negro; two souls, two thoughts, two unreconciled strivings; two warring ideals in one dark body.' (12) Dubois (2013) believed that this psychological disunity was a lamentable consequence of slavery's suppression of autonomous subjectivity. He saw his task as assisting the 'Negro' to attain the consciousness of which he had been forcibly deprived.

G.W.F. Hegel's *Phenomenology of Mind* gave an analysis of slavery that worked on the historical and psychological levels simultaneously. He described a process by which the slave learns to recognize that the master is dependent on him, rather than *vice versa,* as he grows to understand his objective circumstances as the product of his own labor. The slave, defined as one whose activity is alienated from his self, is liberated when he 'realizes that it is precisely in his work wherein he seemed to have only an alienated existence that he acquires a mind of his own.' (cit. Rees 2005 35) Within the individual this liberation consists in the recognition of a subjective consciousness that transcends the body; historically, Hegel's reflections were inspired by the successful slave revolt in Haiti (which he seems, like many others, to have assumed would be the first of many).

In *Beyond Good and Evil,* Nietzsche described the slave as 'an incomplete human being,' lacking the unified subjectivity of the master. However, he viewed modernity as the site of a 'slave revolt in morals,' whereby the conflict between slave and master was internalized, so that the individual mind became 'a battleground for these oppositions.' Nietzsche believed that the modern world would favor the servile mentality in its battle against the 'master morality' with its unified consciousness and autonomous subjectivity, and he predicted the imminent demise of the 'doer behind the deed.'[24] The Darwinian conglomeration of self-seeking genes that some of today's neuroscientists suggest is the real human self is Nietzschean man incarnate. So is the disparate, material subject described by postmodernists like Judith Butler and post-humanists like Donna Haraway. Both represent the apparent psychological and political victory of the slave's psychology over the master's.

That is not, of course, the same as the victory of the slave over the master. On the contrary, the fact that slave consciousness is becoming universal suggests that humanity is learning to love its servitude. But to what are we enslaved? If, as the thinkers discussed above suggest, the struggle between master and slave is now being fought on the internal battlefield of the individual psyche, how can we identify and distinguish the combatants? In every era, theorists of slavery identified it with the tendency towards reification: transformation into a thing. In what practical, demonstrable manner do the human beings of the third millennium perform this operation on themselves?

We have seen how wage labor imposes an exchange-value on human life. In the symbolic guise of money, man's own activity rises up before him in objective form. As it does so, it comes to appear as a subjective actor in its own right. With the incremental relaxation of the prohibition against usury, money seemed to come alive, developing the capacity for autonomous reproduction, learning how to grow and develop in the absence of human intervention. In short, money has become a performative sign. The autonomous power of signs is not an ontological but an historical phenomenon, with identifiable historical causes which become clear once we recognize it as manifested in financial as well as linguistic representation.

The history of money's development displays the exponentially-increasing, autonomous power of representation. For thousands of years, financial value was assumed to be incarnate in precious metals. It was not until the seventeenth century that money began to take the form of paper that was worthless in itself, thus clearly differentiating financial value from its sign. Yet the fetishistic illusion that there was, somewhere, a physical body of gold to which all money ultimately referred officially persisted until the 1970s. The abandonment of the gold standard at Bretton Woods finally dispensed with that fantasy. Money was acknowledged as what it had always really been: an image. Since the 1970s, this image has grown ever more autonomous, and ever less referential, to the point that it is now entirely unconnected to the physical dimension. In the process it has acquired an efficacious power: money is a sign that gets things done.

Today, the hyper-real rule of signs in the cultural realm is paralleled by the rule of financial signs over the economy. The economy no longer stops at the borders of consciousness, if it ever did. Postmodernity seems to have exhausted the distinction between 'culture' and 'economy,' so that we would do better to speak of a 'dictatorship

[24] The connections between Hegel, Dubois, Nietzsche and Aristotle have been brilliantly discussed by Bull 1998. My account is inspired by his, though my reading of the texts is rather different. In particular, Bull reads Nietzsche 1907 as 'reassert[ing] the unified self of the master morality in order to exclude the multiple selves engendered by the success of slave morality.' I agree that Nietzsche admires the master morality, but I think he believes it is beyond salvation. The future, in Nietzsche's view, belongs to the slave morality.

of representation' that spans every aspect of society, and unites what until recently appeared as separate 'fields' or 'areas' of experience. The concept of the 'economy' was invented by early political economists as a *cordon sanitaire* erected around the market. The economic sphere was originally conceived as an arena in which ordinary moral strictures against avarice and self-seeking were suspended.[25] And the concept of a discrete, semi-detatched 'sphere' or 'field' called the 'economy' remained plausible enough through the industrial era, when the visible mechanisms of material production seemed to provide it with a physical as well as a conceptual boundary.

In the age of credit cards and electronic money, however, it seems impossible any longer to draw a meaningful distinction between financial and semantic modes of representation. Financial value in general is revealed as nothing more than figuration.[26] The current economic crisis results from a yet further autonomy of representation within the financial system itself, as financial instruments become self-referential through the chain of derivatives. With regard to human identity ('psychology' no longer seems an appropriate term), the loss of essence gives rise to the manifold forms of 'post-human' culture: artificial intelligence, genetic engineering, cyborgs, prostheses, cosmetic surgery, smart drugs, sex changes, even materialist and linguistic determinism themselves, combine to confirm what is now a virtually *a priori* assumption among the Western intelligentsia that appearance and essence are indistinguishable, if not identical.

Today it is no exaggeration to say that money rules the world. It is no longer plausible to blame money's influence on a malevolent group of human beings such as Marx's bourgeoisie. Money pursues its own interests, even when they contradict the interests of those who nominally possess it. Money determines the way people think. And yet money is nothing but a thought itself. Financial value exists only in the human mind. If we find that our lives and thoughts are increasingly dominated by money, if we find that money determines both international politics and individual personalities, if we conclude that money has achieved a position of domination over the people whose activity it represents, we must interpret these developments as psychological in nature. Money has not conquered the world by physical force. It has conquered the only territory on which it can possibly operate, since it is the only place it can exist. The victory of money has taken place within the mind.

What is money? It is a certain way of conceiving ourselves. It is what we get when we think of ourselves as things. It is human activity imagined as alien to the people who carried it out, separated from them and acting independently of them. It possesses the major characteristics of life: the ability to reproduce, a canny knack of noticing and pursuing the optimal means of ensuring its continued reproduction, mobility, inventiveness and something that looks like aggression. It can and frequently does take material shapes, though these are not constrained by any law of nature. But

---

[25] See Joyce Oldham Appleby 1978, *Economic Thought and Ideology in Seventeenth-century England.*

[26] See Grant, Ferguson C., and Ferguson N, *Money of the Mind: Borrowing and Lending in America from the Civil War to Michael Milken,* Farrar, Strauss and Giroux 1995, and Neill Ferguson 2009, *The Ascent of Money*. For a recent account of autonomous representation's role in causing the financial crises of the early twenty-first century, see Charles Ferguson 2003, *Predator Nation.*

money is not alive. Money is not a conscious subject. Money is merely a symbol of dead labor-power, reified human activity. As the representative of dead human labor, money naturally behaves in ways that resemble those of living humanity, but without any awareness of what it is doing. What is money but a philosophical zombie?

The current popularity of the zombie metaphor is not limited to neuroscientists. It has an obviously apt valence in critiques of postmodern capitalism. Chris Harman's *Zombie Capitalism* recalls how:

> Faced with the financial crisis that began in 2007, some economic commentators did begin to talk of 'zombie banks'—financial institutions that were in the 'undead state' and incapable of fulfilling any positive function, but representing a threat to everything else. What they do not recognize is that 21st century capitalism as a whole is a zombie system, seemingly dead when it comes to achieving human goals and responding to human feelings, but capable of sudden spurts of activity that cause chaos all around.[27]

Anthropologists tell us that societies newly introduced to a money-based economy rapidly develop an intricate network of magical beliefs, often resulting in large-scale witch-hunts. This happened at the dawn of capitalism in Western Europe, and it is happening today in parts of West Africa and South America that are being drawn into the global marketplace for the first time.[28] The idea of a performative sign—an object endowed with fetishistic significance that allows it to alter the condition of the objective world, to do things, such as reproduce, without any human intervention—is after all a magical belief. The magician believes in the efficacy of amulets and talismans, the financier believes in the power of derivatives and hedges. Both put their faith in the power of performative representation.

One of the most common social reactions to the new imposition of alienated labor is the development of zombie myths. These sprang up in West Africa and the Caribbean along with the slave trade. They are clear enough expressions of anxiety and coded protests against the separation of subjective activity from subjective volition that both slavery and wage labor involve. The Africans recently studied by Peter Geschiere speak of:

> … a new type of witches who no longer eat their victims, but who transform them into a kind of zombie and put them to work on "invisible plantations." The *nouveaux riches* supposedly owe their success to the exploitation of these zombies' labor. People insist on the novelty of this form of witchcraft and often relate it to the arrival of the Europeans and the introduction of new luxury items.[29]

Vincent Brown observes that in their response to the development of the slave trade:

> Africans undoubtedly fixed upon the association of slavery and death. Their first assessments of Europeans, their experience of the impact of commercial wealth on the coast, and their impressions and myths about the Atlantic economy reveal a way of seeing, speaking and thing that associated Atlantic slavery with murder, sorcery, and the alienated dead.[30] (38)

[27] Harman 2010, 10–11.

[28] See especially Taussig 1993, van Binsbergen 2001, Englund 1996, and Meyer 1999.

[29] Geschiere 1997.

[30] Brown 2008.

While Edwin Ardener describes the sudden emergence of a zombie cult among the Bakweri of Cameroon after the introduction of money into their society in the mid-twentieth century:

> … a rumor spread that the elders had ordered that no money should be picked up from the ground, since it was being scattered as a lure to entice men to the waterside. There, "Frenchmen" would use them to work as zombies on a new deep-sea harbor, or use them to appease the water-spirits. For a number of months it was commonplace to see coins and even low-value notes lying about the streets of the capital.[31]

The current craze for zombie literature and movies in twenty-first century Western culture is no less symptomatic.[32] In fact the vogue for zombie 'mash-ups,' in which the classics of Western literature are re-written in order to feature a battle against zombies would seem to reflect a yet more profound concern that the past as well as the present is under threat. It reflects a horrified fascination with what Coleridge, in one of the earliest works of zombie fiction, called 'the nightmare Life-in-Death.'[33] But belief in zombies is no more superstitious than belief in the stock market. It is arguable that magic did not disappear from Western societies because it was defeated, but because it was so completely victorious as to render itself invisible. To a rational mind, of course, the magical power of signs and tokens is nothing more mysterious than human power disguised in alien form. But twenty-first century society is not ruled by rational minds. If the conscious self truly is obsolete, we will not be ruled by such minds again. Are we to be ruled by zombies instead?

References

Althusser L (1971) Ideology and ideological state apparatuses. In: Lenin and philosophy and other essays (trans: Brewster B). New Left Books, London, pp 160–161

Appleby JO (1978) Economic thought and ideology in seventeenth-century England. Princeton University Press, Princeton

Austin JL (1962) How to do things with words. Oxford University Press, Oxford, p 91

Barthes, R (1978) The death of the author. In: Image, music, text (trans: Heath S). Hill and Wang, New York

Behn A (2003) Oroonoko, or the royal slave. Penguin Books, London

Bradley K (1994) Slavery and society at Rome. Cambridge University Press, Cambridge

Bradley K (2000) Animalizing the slave: the truth of fiction. J Rom Stud 90:110–125

Brown V (2008) The Reaper's garden: death and power in the world of Atlantic slavery. Harvard University Press, Cambridge, MA

Buck-Morss S (2000) Hegel and Haiti. Crit Inq 26(4): 821–865

[31] Edwin Ardener, "Witchcraft, Economics, and the Continuity of Belief" in Douglas (1970), 141–160, quotation from 154.

[32] There is a plethora of zombie-criticism around at present. See for example Christie and Lauro 2011.

[33] Samuel Taylor Coleridge, 'The Rime of the Ancient Mariner.' This image has often been interpreted as a reference to the slave trade. See for example, Empson 1964, Ebbatson 1972.

Bull M (1998) Slavery and the multiple self. New Left Rev
Bull M (2000) Seeing things hidden: apocalypse, vision and totality. Verso, London
Chalmers D (1995) Facing up to the problem of consciousness. J Conscious Stud 2(3):200–209
Christie D, Lauro SJ (eds) (2011) Better off dead: the evolution of the zombie as post-human. Fordham University Press, New York
Cicero (1991) On duties. Atkins EM (ed). Cambridge University Press, Cambridge
Cooke B (2003) The imprint of slavery in management studies. J Manag Stud 40:1895–1918
Dennett D (1989) The intentional stance. Bradford Books, Cambridge, Massachusetts
Dennett D (1997) Kinds of minds: towards an understanding of consciousness. Basic Books, New York
Dubois WEB (2013) The souls of black folk. Eucalyptus Press, Oakland, CA
Ebbatson JB (1972) Coleridge's mariner and the rights of man. Stud Romant 11:171–206
Empson W (1964) The ancient mariner. Crit Q 6:298–319
Englund H (1996) Witchcraft, modernity and the person: the morality of accumulation in central Malawi. Crit Anthropol 16(3):257–279
Epictetus (1916) The discourses of Epictetus (trans: Matheson PE). Retrieved from http://classics.mit.edu/Epictetus/discourses.html
Finley M (1973) The ancient economy. University of California Press, Berkeley
Ferguson C (2003) Predator nation: corporate criminals, political corruption and the hijacking of America. Crown Business
Ferguson N (2009) The ascent of money: a financial history of the world. Penguin Books, London
Foucault M (1969) What is an author?. Retrieved from http://foucaultblog.wordpress.com/2007/05/29/full-text-of-what-is-an-author/. 30 April 2012
Geschiere, P (1997) The modernity of Witchcraft: politics and the occult in Postcolonial Africa (trans: Geschiere P, Roitman J). University Press of Virginia, Charlottesville
Glancy J (2002) Slavery in early Christianity. Oxford University Press, New York
Graeber D (2001) Towards an anthropological theory of value. Palgrave, London/New York, pp 412–419
Graeber D (2007) Possibilities: essays on hierarchy, rebellion and desire. AK Press, Oakland
Grant J (1995) The money of the mind: borrowing and lending in America from the civil war to Michael Milkin . Farrar, Strauss and Giroux, New York
Harman C (2010) Zombie capitalism: global crisis and the relevance of Marx. Haymarket Books, Chicago, pp 10–11
Hegel GWF (1967) Philosophy of right (trans: Knox TM). Oxford University Press, London
Hobbes T (2003) Leviathan. Tuck R (ed). Cambridge University Press, p 141
Johnson W (1999) Soul by soul: life inside the antebellum slave market. Harvard University Press, Cambridge, MA, p 64
Lafargue P (1883) The right to be lazy and other studies (trans: Kerr CH). Charles H. Kerr and Co., Chicago, 1907
Marx K (1967) Capital, vol 1. International Publishers, New York, pp 59–60
Marx K (1993) Grundrisse: foundations of the critique of political economy (trans: Nicolaus M). Penguin Books, London
Meltzer M (ed) (1993) Slavery: a world history. Da Capo Press, New York
Milton J (1959) Complete prose works. Wolfe DM (ed). Yale University Press, New Haven
Meyer B (1999) Commodities and the power of prayer: pentecostalist attitudes towards consumption in contemporary Ghana. In: Meyer B, Geschiere P (eds) Globalization and identity: dialectics of flow and closure. Blackwell, Oxford, pp 151–176
Nietzsche FW (1907) Beyond good and evil. Zimmern H (ed). The Macmillan Co., New York
Patterson O (1982) Slavery and social death: a comparative study. Harvard University Press, Cambridge, MA
Plato (1892) The dialogues of Plato: republic, Timaeus, Crito (trans: Jowett B). The Macmillan Co., London
Rees J (2005) The algebra of revolution: the dialectic and the classical Marxist tradition. Routledge, New York

Rodriguez JP (1997) The historical encyclopedia of world slavery (2 vols). ABC-CLIO, Santa Barbara
Shannon J (1849) The philosophy of slavery as identified with the philosophy of human happiness, vol 29. A.G. Hodges & Co, Frankfort
Smith A (1976) An inquiry into the nature and causes of the wealth of nations, The works and correspondence of Adam Smith. Campbell RH, Skinner AS (eds). Glasgow
Taussig M (1993) The devil and commodity in South America. In: Comaroff J, Comaroff J (eds) Modernity and its malcontents: ritual and power in postcolonial Africa. University of Chicago Press, Chicago
van Binsbergen W (2001) Witchcraft in modern Africa as virtualized boundary of the kinship system. In: Bond GC, Ciewkawy DM (eds) Witchcraft dialogues. Ohio University Press, Athens, pp 212–264
Watson A (ed) (1985) The digest of Justinian. University of Pennsylvania Press, Philadelphia

Chapter 18
Mind and Brain: Toward an Understanding of Dualism

Kristopher G. Phillips, Alan Beretta, and Harry A. Whitaker

18.1 The Views of Antiquity

Starting two and half millennia ago in Greek antiquity and continuing to twenty-first century paradoxes, the story of dualism – the mind-brain problem – has many variations but retains some common themes. One of the earliest natural-philosophical investigations of the mind can be found in Homer's Iliad and Odyssey, epic poems thought to have been written in the seventh or eighth century BCE (Green and Groff 2003). While one assumes that Homer's main intention was to write heroic poetry and not to describe a psychological theory, his work nevertheless contains some interesting insights into the precursors of Western theories of mind. First, the Homeric Greeks appear not to have had a cohesive conception of mind nor of body (Jaynes 1990); there are no terms in Homer for words comparable to "soul" or "mind" or "thinking" or "perceiving" (Russo and Simon 1968). Instead in the Iliad and Odyssey we find terms such as *psyche*, *thymos*, *menos*, *phrenes*, and *ate* (Green and Groff 2003).

The word *psyche* appears in the first sentence of the Iliad. In both the Iliad and the Odyssey, *psyche* is compared to a "phantasm" or *eidolon* (Sandywell 1996). Like a phantasm, *psyches* are capable of haunting people after they have departed the body, as Patroclus haunts Achilles until Achilles will properly care for Patroclus' body. The *psyche* is ostensibly useless while inhabiting a living body; it does not appear to engage in cognition, emotion, or will. The *psyche* is the life-force of a thing and leaves the body when the body dies (Green and Groff 2003). After death,

K.G. Phillips (✉)
Department of Philosophy, Northern Michigan University, Marquette, MI, USA
e-mail: kristopher.g.phillips@gmail.com

A. Beretta
Department of Linguistics, Michigan State University, East Lansing, MI, USA

H.A. Whitaker
Department of Psychology, Northern Michigan University, Marquette, MI, USA

C.U.M. Smith and H. Whitaker (eds.), *Brain, Mind and Consciousness in the History of Neuroscience*, History, Philosophy and Theory of the Life Sciences 6,
DOI 10.1007/978-94-017-8774-1_18,

the vaporous, dispassionate *psyche* is exhaled and Hermes escorts it to Hades (Sandywell 1996). Two terms the ancient Greeks used to refer to what we now consider to be psychological processes are *thymos* and *nous*. *Thymos*, or *thumos*, is the source of emotion, sometimes with a quasi-intellectual aspect, and is strongly connected to motivation and will; its physical source is the diaphragm (Green and Groff 2003). The term *thymos* can be used to mean a particular impulse itself or, non-dualistically, the organ from which the impulse originated (Russo and Simon 1968). *Thymos* departs the body upon death, but does not go to Hades as does the *psyche*. Instead it appears to dissipate. Unlike the *psyche*, non-human animals are said to possess *thymos* (Green and Groff 2003).

The *noos*, or *nous*, is far more intellectual than the *thymos*. It was said to be located in the chest. The *noos* is somewhat similar to the mind or a product of the mind; it is not mentioned often in Homer's work, likely because his heroic poetry was more concerned with warriors than thinkers. Although the term *noos* cannot be translated literally as "the mind" it can refer to the organ of planning or to the capacity one has to plan or even to the plan itself. Avoiding or perhaps simply uninterested in a dualistic account, Homer does not distinguish between these uses; which one is intended at any given point is left for the reader to discern from context (Russo and Simon 1968).

Two other Homeric psychological terms are *menos* and *ate*. *Menos* is similar to *thymos*, but carries with it a more narrow range of application: *menos* is a divinely caused surge of battle prowess for a warrior who is on the verge of giving in to fatigue or despair on the battlefield, while *thymos* applies to the source of emotions generally. *Ate* is a kind of madness, and, like *menos*, often has a divine cause. For instance, when Agamemnon eventually returns the slave girl to Achilles, Agamemnon says his action comes from an *ate* caused by Zeus. In Homer's time, attributing mental states to external forces such as the gods, was not uncommon. It was not until sometime after Homer (perhaps 650–600 BC) that people began to write as though they were self-responsible for their mental states (Green and Groff 2003).

It is interesting to note that in the Homeric epics, it is possible for a person to converse with the *thymos* or *phrenes* almost as one would speak with another person. In the Odyssey, Odysseus engages in what might first appear to a modern reader to be a soliloquy about the perils of being lost at sea. As the apparent soliloquy continues, it becomes clear that Odysseus is not speaking to himself, but is instead engaging in a conversation with his *thymos* (Russo and Simon 1968). One could argue that a character in a Homeric epic does not engage in any activity in isolation; he is always in the presence of his *thymos*, *menos*, *ate*, and so on. An Homeric character's actions are not entirely his own – his hand can be stayed or his mind changed by the intervention of a god. Keeping these terms and concepts in mind, we can begin to frame a Homeric model of the mind, defined as follows (Russo and Simon 1968).

- Mental activity is represented as an exchange between personified concepts such as the *thymos*.
- Mental activity can be started by forces external to the person, such as a god or another person.

- There is no obvious distinction between organs, the activity of the organ or the products of the activity of the organ.
- Mental activity is comprehensible and visible.
- The "self" is defined in a series of exchanges with others.

Thales, the earliest of the pre-Socratic thinkers averred, 'All things are full of gods'.[1] Indeed it has been argued that for these early thinkers the entire natural world was alive and vital.[2] How 'mind', 'spirit' or 'subjectivity' was related to matter, let alone the body or brain, was not a problem, it was simply a state of affairs. Yet the way in which they were 'attached' differed substantially between the different schools of pre-Socratic philosophers. One might presume, on the basis of the fact that mind-body interaction was taken as obvious, that the problem of substance dualism hardly intruded into the work of the greatest natural philosopher of the ancient world, Aristotle. Of course, such a suggestion requires some support, and a cursory discussion of other views is in order before we consider the Aristotelian view. Aristotle's approach can and we think, should, be understood largely as a response to the views of those that came before him.[3]

The pre-Socratic view that likely most influenced Aristotle's own views (and happens to be in some ways similar to the views of the seventeenth Century philosopher René Descartes[4]) belongs to the Pythagoreans. A striking departure from the hero myths of the Iliad and the Odyssey, Pythagoras is said to have had his followers study the work of an Asian-influenced Greek philosopher-poet by the name of Pherecydes, who maintained that the fear of death is unfounded since we survive the death of our earthly bodies and experience a rebirth in different bodies. The purpose of life, the Pythagoreans thought, was to prepare the soul for a final separation from material bodies in order to make the journey "home" to the ethereal plane. The nature of the soul, according to the Pythagoreans, was an immortal, immaterial divine entity that was roughly mathematical in nature (see Aristotle, *De Anima* I; Apostle and Gerson 1991). Himself a mathematician, Pythagoras noticed that the truths of geometry were not perfectly manifested in the ever-changing material/sensible world. Indeed, such truths were found only deep in the soul. The sensible world was an imperfect image of the divine world found within the soul—for instance, harmony found in music represented the kind of harmony that the soul was capable of when it finally reached the divine-world.

Pythagorean metaphysics is a complicated business, but for our purposes we can focus on that which made its way to Plato, and from Plato to Aristotle. The kind of

---

[1] Thales, as reported by Aristotle (*De anima*, 411a7); in Kirk and Raven 1971, 94.

[2] Frankfort and Frankfort 1949, 14: 'primitive man simply does not know an inanimate world.... The world appears to primitive man neither inanimate nor empty but redundant with life…'.

[3] See, for example (Apostle 1969), *Metaphysics* A, wherein Aristotle critically surveys the views of his immediate predecessors.

[4] It is worth noting that while the spirit (so to speak) of the Pythagorean view in some ways parallels the spirit of the Cartesian view, there are important differences in both the physics and metaphysics that inform the views.

dualism present in the Pythagorean conception of the world was what we will here call a "form-matter dualism." According to Pythagorean metaphysics, the world and everything in it is made up of various ratios of *peras* (form) and *apeiron* (matter). It is from this dualism of form and matter that we can make sense of the soul's nature. For the Pythagoreans, the soul was made up of "the first *tetractys*" (the numbers, 1, 2, 3 and 4, which together add up to the "perfect number," or 10). The body, then, was composed of the point (1), the line (2), the plane (3) and the solid (4), which are the geometric correlates of the numbers. The soul possessed the powers of intellect, inferential knowledge, belief, and sensation. This seems to be the one of the first times that the soul was equated with what we now take to be the mind[5] and is a view that Descartes will champion much later. What should be noticed is that despite the importance of the *separability* of the soul from the body, their interaction is taken as a basic assumption. That there is a physical world and that it correlates to the mind can be accounted for, according to a Pythagorean, mathematically, as the correlate between ethereal number and its geometric manifestation. For them, this seems to suffice by way of an explanation.

Plato's view regarding the soul can be seen as a systemization of the Pythagorean form-matter dualism, incorporating the contributions of Heraclitus and Parmenides (Duerlinger 2005). Plato's account is notably less mathematical in nature than his Pythagorean predecessors, but the account of soul is strikingly similar. The soul itself is immortal, and this is guaranteed by the very essence of it (see Plato's *Phaedo*; later Aristotle defends Plato's approach to proving the immortality of the soul in *Posterior Analytics*). The soul, Plato famously argues, has three different parts: the intellectual, the spiritual, and the appetitive (*Republic* IV) (Hamilton and Cairns 1961). Each part plays an important role in giving life to the body. The appetitive keeps the body alive but tends to be somewhat unruly; the spiritual seeks honor, glory, and is the source of the passions; the intellectual is charged with the role of keeping the other parts of the soul in check – it does all of the heavy lifting, so to speak, with regard to reason-based activity, reflection, and so on. Plato suggests in the *Phaedo* that the soul is most like the forms—that is, ethereal, timeless, immortal, and the object of true wisdom. In Plato's *Timeaus* it is argued that "tiny nails" connect the body and soul, a view that is reiterated, albeit metaphorically, in the *Phaedo* (Hamilton and Cairns 1961). Still, such an account of the fastening of the soul to the body fails to provide an illuminating answer to how the form and matter interact.

In *Physics* I 9, Aristotle attributes the view to Plato that mind and body are *essentially* different, and thus *cannot* interact (Apostle 1966). Furthermore, in *Metaphysics* A, Aristotle criticizes Plato for never offering an account of how the forms are able to interact with the sensible world (Apostle 1969). This point is important for several reasons. First, we see here the introduction of the notion of the essence of a thing. Essentialism plays an important role in Aristotelianism and the Scholasticism that dominated Western thought for over a thousand years after Aristotle; some notion

[5] Some scholars maintain that the accompanying creation doctrine is what Plato presents in his dialogue *Timaeus*. Others, Duerlinger (2005) e.g., think that Plato himself adopted a view very similar to this Pythagorean view.

of essentialism even plays a role in Cartesian arguments for substance dualism. Second, understanding form-matter dualism is crucial for Aristotle's hybrid account, wherein Aristotle collapses the Platonic parallel form and matter worlds into one. Finally, this parallelism of form and matter might be seen as a precursor to Leibniz's famous doctrine of pre-established harmony—wherein the mind and body never actually interact; the universe is merely set up in such a way that the order of mental events directly parallels the order of the physical thus making it *appear* as if physical and non-physical beings can interact.

All of these are more spiritual and/or philosophical accounts of the mind than they are scientific (although the distinction as we know it now between philosophy and science is a rather new one, relatively speaking). In Aristotle's work we find a view that, while still very far from current science, more closely squares with what we take to be the dominant approach today; his naturalistic approach is at least much closer to what we understand as scientific than those views that incorporate separate insensible worlds and the doctrine of rebirth. Indeed, "If, then, one means by "identity theory" simply the contention that what we call mind and its manifestations are not separable from the body, surely Aristotle… subscribed to this theory" (Matson 1966, 93). Although Aristotle was Plato's student and thus inherited much of Plato's metaphysical baggage, an important difference between the two is that, while Plato did not seem to think that form and matter could interact (or at least did not have a plausible account of such interaction), Aristotle thought that all they did was interact. In fact, in Aristotelian terms, form cannot exist without matter, and matter cannot exist without form.[6] For Aristotle, the form provides the essence, or the nature of the thing, while the matter provides the physical stuff to be made into that thing. For example, what makes a tree an oak tree rather than a cow is that the physical stuff that makes up the tree possesses the form of an oak tree. We can come to know the features or characteristics of that tree by coming to know the features that define what it is to be an oak tree.

This is relevant to Aristotle's notion of the human soul. What kind of form that matter takes on will directly impact the nature of the mind present in the being who exhibits that form. Substantial forms, as Aristotle calls them, are the soul, and different kinds of souls possess different parts, depending on what kind of form it is. There were three kinds of souls, those of plants, those of all animals including us and those specifically reserved for humans. In many cases the form does not provide for a soul, and thus there is simply no mind. Rocks, for example, do not have minds in the sense that we now think about it (although Aristotle does talk about rocks "wanting" to return to the earth, or fire "trying" to make it back to the sun because these are their natural places). Like those before him, Aristotle maintained that the soul was what gave life to a creature. And like Plato before him, Aristotle thought that the soul was divided into parts. Depending on the kind of soul a being possesses, it may have some or all of the following parts: [1] the reproductive/nutritive part of the soul (present in all living beings) which serves to reproduce more beings like it as well as to

[6] There is one notable exception to this general rule: God. According to Aristotle, the divine being is form itself, without matter. See, *Metaphysics* Γ, Z, H and Λ.

nourish the body, [2] the sensitive-imaginative part of the soul (present only in animals) is physically located in the sense organs of a living thing allowing for a creature to take in information about the world around it and to react. Do not let the term “imaginative” fool you, however; this refers not to flights of fancy, but rather to the ability to reproduce mental images of those sensory experiences taken in the present or past. [3] Next is the desiderative-locomotive part of the soul. This part of the soul (present only in animals, including human animals) gives beings the ability to desire pleasurable sensations, and to seek to avoid those sensations that cause discomfort; it also allows for the being to move around in an attempt to pursue or avoid such stimuli. [4] Finally, the intellective part of the soul (which exists only in humans) serves as the power of the soul to engage in rational reflection. The Aristotelian conception of the soul is rather directly reflected in the first systematic efforts to establish these functions in the brain, medieval cell doctrine (see Whitaker 2007).

For Aristotle, even though the soul has different parts, it is numerically single, albeit spread out throughout the body. The sensitive portions of the soul actually exist *in* the sense organs and as such, it may appear that Aristotle is the closest of the Greeks to modern-day physicalist accounts of the nature of the mind.[7] However, such an assumption is not warranted. Aristotelian metaphysics are weird, and despite the actual physical location of the mind/soul in the body, there is still an important kind of dualism present in Aristotle’s work. At the very least, there remains a form-matter dualism (nearly all things are composites of form and matter); there is still an important distinction between the material that makes a thing up (the material cause), and that which makes it what it is *essentially* (the formal cause). So despite Aristotle’s proclivity for naturalism, and the collapsing of the patently dualistic world-view of Plato and the Pythagoreans, there still exists an important dualism in Aristotelian natural philosophy.

Still, the dualism present in the Aristotelian system is not *obviously* the kind of dualism philosophy professors teach to introductory students today. The kind of dualism taught in philosophy courses today more closely resembles developments derived from Renaissance philosophers (notably Cartesian substance dualism, or the more recently developed property dualism). Matson (1966) has argued that such a conception of dualism is entirely foreign to all the Greeks, and that maybe they (the Greeks) had the right idea. If Matson is correct, there is something instructive about the *kind* of dualism present in Aristotle’s natural philosophy, even though that dualism is radically different from how we think of it today. If Aristotle and his Greek peers never considered the mind-body problem (or soul-body problem) in the way that we do today, it is not hard to see how something like his conception of the soul can lead to the conception we do grapple with today.

The break-up of classical civilization in the fifth century CE that ushered in the so-called ‘dark ages’ into most of Europe was shortly followed by the emergence of a brilliant civilization in the Islamic Middle East. Although a great deal of the Islamic ‘golden age’ was based on Arabic translations of the major works of Greek

[7] For a defense of the view that Aristotle did not consider the body and soul separate in any sense, see Matson 1966.

antiquity, it also saw an independent development of early experimental science. One thinks of Ibn al-Haytham's (known as Alhazen in the west) optics[8] as well as labor-saving technologies, especially those powered by water. With the post-Crusades' translation into Latin of the major part of Arabic learning in the twelfth century *et seq*, Arabic developments in natural philosophy, medicine and technology were (re-) introduced into Europe. Concomitantly, during the later Middle Ages a reorientation away from the things of the next world towards the things of this world, slowly accelerated. Technologies based on human and animal muscle power slowly gave way to technologies based on wind and water power.[9] The concept of the inorganic, of efficient rather than final causes, increasingly took hold.

18.2 The Medieval Perspective

Depending on our personal convictions, today we may consider mind to be something distinct from soul (of course not everyone would speculate on either the existence or the contents of the soul). But for the medieval and Renaissance thinkers, mind and soul were intertwined and man's soul was divided into parts roughly following the Aristotelian model, the Platonic model or a creative alternative. The diagram makers who represented medieval cell doctrine sometimes represented a three-part soul (Dryander, Reisch-Brunschwig) and sometimes a five-part soul (a translation of Avicenna); sometimes the parts were divided in two, creating a six-component model (the 1503 Bologna drawing, Albertus Magnus) and sometimes each part represented a single function (see Whitaker (2007) for discussion). What concerns us here is the rational part of the soul: it comprises both of the other kinds of soul, but it adds reason and is not shared with plants or other animals. Rational soul, in all of its complex interpretations, is the nearest anyone came to the contemporary notion of mind (see Lewis 1964, pp. 153–165).

Medieval scholars recognized that there needed to be some means for (rational) souls, which were considered immaterial, to interact with bodies. And they were not shy about proposing solutions. The picturesque solution presented by Alanus ab Insulis, a twelfth century French thinker and poet, was that "the soul is fastened to the body *gumphus subtilibus*, 'with tiny little nails'" (cited in Lewis, p. 60 as from Plato's *Timaeus*, 43a).

It was commonly believed that some sort of 'subtle gumphus' was required to bind body and soul together and if it was not going to be 'tiny little nails', then it had to be some other entity. As Lewis, put it, "this *tertium quid*, this phantom liaison-officer between body and soul, was called *Spirit* or (more often) the *Spirits*… The spirits were supposed to be just sufficiently material for them to act upon the body, but so very fine and attenuated that they could be acted upon by the wholly immaterial soul" (p. 166–7). For example, in Timothy Bright's sixteenth century *Treatise of*

---

8 Al-Haytham's *Optics* was published about 1025 CE; see Sabra (1983, 2002).

9 White 1963. White gives a number of reasons for the development of labour-saving technologies in medieval Europe, not least 'the spiritual repugnance of subjecting anyone to drudgery' (p. 291).

Melancholy (Bright 1586), the spirits are, "a true love knot to couple heaven and earth together; yea, a more divine nature than the heavens with a base clod of earth", so that the soul is "not fettered with the bodie, as certaine Philosophers have taken it, but handfasted therewith by that golden claspe of the spirit" (Lewis, p. 167).

The other attempt to ease the ever-mounting worries about how the mind and body interact is the notion of 'animal spirits' which, as we've noted, was the medieval cell doctrine solution, particularly popular from the fourteenth to the sixteenth centuries; this was the model that René Descartes chose to use. Lewis, a scholar of medieval literature and thought who rejoices in many of the peculiarities of the medieval worldview, emphatically draws the line here: "This doctrine of the spirits seems to me the least reputable feature in the Medieval Model. If the *tertium quid* is matter at all (…) both ends of the bridge rest on one side of the chasm; if not, both rest on the other" (p. 167).

The difficulties with substance dualism slowly came into focus during this period. At least from the time of Galen and possibly before, the status of these 'animal spirits' present in the brain's ventricles and (most believed and Descartes famously drew) in the nerves and blood vessels, was to say the least, ambiguous. Were they merely messengers from the 'soul' lodged in the 'marrow' of the brain or did they possess psychic qualities of their own? Vesalius, in the mid-sixteenth century, began to suspect that the watery fluid in found in the ventricles was no more than, precisely, a watery fluid, and speculated (in print at least) no further.[10] It was only with Descartes in the seventeenth century, inspired by the Copernican/Galilean revolution in natural philosophy toward which the long centuries of the medieval period had been headed, that a final and complete separation of mind from body occurred. Descartes' *L'Homme* (Cottingham et al. 1984) treated the animal as an 'earthen machine', animal spirits as no more than the spirits one might find sold in any bar, while mind became, in Gilbert Ryle's phrasing, 'a ghost in the machine' (Ryle 2002).

The medieval view of the mind-body problem is distinctly unsatisfying to the modern way of thinking, notwithstanding its representation of a componential mind with distinct geographic locations in the brain. With that background in mind, let us now turn to Descartes and his sharply drawn distinction between mind and body, without the complication of a tripartite soul and the mediation of a range of spirits, but crucially, with a clear guiding principle regarding the nature of 'body'. This clarity will make it far easier for us to see why the Cartesian formulation of the mind-body problem is not the problem facing modern scientific inquiry.

18.3 The Cartesian Formulation

Cartesian dualism is a rather clear concept: Mind and soul are one and the same substance, it is immaterial, and the essence of it is thought. Indeed, in a departure from the Aristotelian conception, Descartes denied that, properly understood, the "soul" could exist in anything other than a human. He wrote, "I do not admit that

[10] Vesalius 1998–2009, Book VII, Chapter 1, p. 624.

the powers of growth and sensation in animals deserve the name 'soul', as does the mind in human beings. This common view is based on ignorance of the fact that animals lack a mind" (letter to Regius, May, 1641) (Cottingham et al. 1991). On the other hand, physical bodies are a completely different substance, the essence of which is mere extension. Neither substance requires the other to exist (the mind can and eventually will exist entirely without physical body and vice versa). Cartesian physics, following the Aristotelian dictum that "nature abhors a vacuum," maintained that the universe is a plenum, and as a result of this, only mechanical forces can and do move physical bodies. Perplexingly, on the Cartesian model the mind has the ability to move the body without apparent direct mechanical contact. Descartes famously 'solved' this problem by proposing that the pineal gland is the gateway from the body to the soul (and the soul to the body).

Descartes knew what *body* (or *the physical*, or *matter*, or *material substance*) was. It was *res extensa*, stuff that occupied, or had extension in space. He also knew how motion of such stuff could be effected. Motion had to be effected by direct contact with other stuff that occupied space. These two key properties of bodies sustained his mechanical view of the natural world; nature could be understood in just the way we might understand how a machine works. Just as Descartes' contemporaries could appreciate the action of springs, cogs, and wheels that bring about the motion of the hands of a clock, so they could appreciate that the planets are pushed around the sun by a celestial fluid, and so they could appreciate all natural phenomena. It is worth noting that the celestial fluid was no different *in kind* from any other physical stuff. All of the stuff that existed in the natural world had the same essence, extension.

All except mind. The essence of mind, Descartes maintained, was thought alone. From this, Descartes seemed to know that it did not occupy space and thus could not be subject to mechanical laws. Mind could plainly bring about motion (I decide to move a finger, and I do), but it could not do so by contact mechanics. It is this characterization that makes most clear the mind-body problem. The contrast between body and mind in this formulation is thus sharply drawn. Body is everything extended, and everything behaves the way a machine behaves, everything except mind. So, mind-stuff and body-stuff must be distinct substances. This sharp distinction is what gives the Cartesian mind-body problem its crispness. Its wide acceptance is in large part due to its clarity, no doubt, but also to its consonance with our commonsense intuitions. The world was just the way it seemed to be; it had objects in it, bodies that moved in ways prescribed by the laws of mechanism, that is, via direct contact.

But the clarity of the distinction between mind and body is a dual-edged sword. The clarity of the distinction and the epistemological argument for the distinction (in Descartes' Sixth Meditation) provided a clear challenge. Princess Elisabeth of Bohemia articulates the worry that becomes apparent when considering Cartesian mind-body interaction most acutely in her correspondence with Descartes. It is worth quoting in full. She says,

> I beseech you tell me how the soul of a man (since it is but a thinking substance) can determine the spirits of the body to produce voluntary actions. For it seems every determination of movement happens from an impulsion of the thing moved, according to the manner in which it is pushed by that which moves it, or else, depends on the qualification

> and figure of the superficies of the latter. Contact is required for the first two conditions, and extension for the third. You entirely exclude extension from your notion of the soul, and contact seems to me incompatible with an immaterial thing. That is why I ask of you a definition of the soul more particular than in your Metaphysic—that is to say, for a definition of the substance separate from its action, thought (Letter to Descartes, 6 May, 1643, in Atherton (1994)).

Mind should not be able to move anything physical and yet it does. There can be no 'tiny little nails' to solve the problem, no superfine spirits, no hocus-pocus of the Medieval sort, but a mystery that is a consequence of a mechanical view of the world. Descartes never really had a clear or satisfactory answer for Elisabeth. Famously, he maintained that the soul is connected to the body at the pineal gland.[11] How exactly it is that the soul interacts with the pineal gland Descartes tried to explain to Elisabeth, and plenty of scholars have since tried to make sense of his response to her. Unfortunately, like Plato before him, the best he could come up with was something of a hand-waving analogy and references to simple and ultimately unsatisfactory ideas.

Newton's gravitational force shattered the mechanical worldview. A force does not occupy space and it moves objects without direct contact. So, the notion of body that stood in clear contrast to mind in the Cartesian formulation of the mind-body problem faced a new challenge with this new physical theory. According to the Cartesian Programme, the mind was of a fundamentally different kind than bodies. Despite the difficulty of reconciling *how* the mind interacted with the body, it was clear that it did. We could examine and understand mind and body separately; each one existed in a different domain of inquiry. When forces were introduced into physical theories, the hard and fast distinction between mind and matter was lost, and with it, its usefulness. In the new physics, body could be *res extensa* or not, and so the contrast with mind, which previously was the only thing that was not *res extensa*, had been consigned to oblivion. *Res extensa* itself was not possible without the basic forces. As Priestley observed: take away the forces, and there would be no solid objects, not even an atom, as everything would be dispersed and there would be nothing left for the imagination to fix upon.

The Cartesian mind-body problem was deflated because 'body' had ceased to have the character upon which the problem was premised (it was also challenged on quite different grounds by La Mettrie (1748/1996) and Locke (1690/1975)). Indeed, since the new science was dealing in forces, vacuums, and other kinds of things that were 'impossible' given the Cartesian world-view, there was no means of maintaining the distinction between material and immaterial 'substances.' That is, built into the new 'physical' science were theoretical entities that did not occupy space and could effect movement across a vacuum without direct contact, eliminating any clear divide between mind and body. To put the point succinctly, the realm of what counted as physical, and thus within the purview of the physical sciences (or "natural philosophy") was anything that entered theory and that evidence indicated was real.

[11] See, for example, Letter to Meyssonnier, 29 January, 1640; to Mersenne, 1 April, 1640; and to Mersenne, 30 July, 1640.

18.4 What Does Mind-Body Dualism Mean Now?

It's a matter of some perplexity why the expression *Cartesian dualism* continues to have the sort of currency that suggests that the Newtonian revolution never occurred and that we are still basking in the psychological luxury of nature appearing to be just like our common sense dictates it should be. The once principled and clear divide between mind and body lost much of its justification once the consequences of forces were taken on board. So, it does not seem reasonable that anyone today would subscribe to Cartesian dualism (where the view is understood as a *scientific* hypothesis regarding the *essential* and complete distinction between mind and body). However, the mind-brain problem is very much alive and well; that it remains a serious and interesting problem is well articulated in the delightful book *Mind, Brain and the Quantum: the Compound "I"* by Michael Lockwood (1989).

And yet, although we plainly have no reason to create a physical-nonphysical 'substance' divide, many proceed as if such a divide obviously exists. To illustrate this, we will take some examples from highly accomplished scholars who care deeply about language-brain relations. Pulvermüller (2002), for instance, is concerned about the failure, as he sees it, of linguistics to translate itself into a plainly physical account, one based on neurons. Instead, recalcitrant linguists appear to think that "language theories must be formulated in an abstract manner, not in terms of neuron circuits", excluding "explicit reference to the organic basis" of language; "for a scientist this may be difficult to understand". This incomprehensible "linguistic mentality" is comparable to that of a "scholar who studies stars but refuses to speak about their component substances and driving forces". What is called for are "translations between the language of linguistic algorithms and that of nerve cells"; "a language theory at the neuronal level is required", but linguists continue to neglect this need; "it is not enough to provide abstract descriptions of language phenomena; it is also necessary to spell out possible language mechanisms in terms of neuronal circuitry" (270–3). On Pulvermüller's account, linguistics has failed to reduce itself to purely physical matter. However, his allegations assume a qualitatively unwarranted divide. On the physical side, he does not question the failure to translate neural accounts into the structures and structural relations of linguistic theory; however, on the abstract side, he proclaims the failure to express theory in physical (neural) terms. Apparently, the physical side is more real or at the very least, has some privileged status that the abstract side lacks. We see the same view in many others who think about how the language-brain problem can be resolved. To cite another recent example, "language researchers who fail to embrace biological approaches will be increasingly left behind" (Margoliash and Nusbaum 2009, 510). Left behind! As the seventeenth century poet Henry Vaughn put it "They are all gone into the world of light / And [linguists] alone sit lingering here." What seems to worry these authors is that core linguistic proposals are made unconstrained by developments in neuroscience ("Neurobioogy and Neuroethology" op cit, 505). Evidently, it never occurred to them that they might equally register angst that core neurobiological proposals are made unconstrained by developments in linguistics. One side of the divide has been conferred a special status; this is what Noam Chomsky referred to as *methodological dualism*.

Yet we have seen that since Newton, such a simplistic divide has no justification. The 'physical', whatever that is, is not more real and has no privileged status at all; indeed, as argued in Beretta (2014) it has *no* useful status. All we have are phenomena in the world that present themselves to us, and we try to understand them as best we can; that is, we construct theories about them and test them. To the extent that these theories are currently not integrated with each other, we hope that one day our understanding will be such that integration will be possible.[12] But the physical-supremacy approach, a standard way of viewing the problem of language-brain relations, has its basis in a discarded model. The demand for physical explanation, and the attendant frustrations of those who divide the world up that way, as Pulvermüller, Margoliash, Nusbaum and many others continue to do, has many historical parallels, as we have seen (Beretta 2014).

In the light of the history of natural inquiry over the last few centuries, what are we to make of such concerns as Pulvermüller's? Recall his view that linguistics is unacceptable for devising its theories in an abstract manner, by which he means 'not in terms of neuron circuits', an approach that he judges would be difficult for a scientist to understand. However, physics, by all accounts, is a ghost field (as Kline (1985) puts it); that is, physics is so abstract that 'physical' has no useful meaning; linguistics is abstract, another ghost field, and so is neurobiology. Which is only to say what has become so abundantly clear since Newton that it is now commonplace: inquiry is governed by theories, however abstract, bizarre, absurd, etc., they may seem to be, so long as they agree with experiment, in other words, with our best attempts to acquire an understanding of aspects of nature. Thus, to assume that theory in one domain, which is allegedly *not* abstract, in some sense yet to be explained, has a privileged status over theory in another domain, requires justification; lacking any to date, the assumption is truly rather difficult to understand, let alone accept.

But perhaps what is meant when many scholars refer to dualism today is that there is a problem that we all face: how to unify our theories of mind and our theories of brain. No one, least of all the present authors, could argue with that. Lest one minimize the problem, however, we do realize that the theory-unification problem is an enormous challenge, the solution which some commentators (e.g., Nagel 1974; Chalmers 1995, 1996)) believe may be beyond us, or "if it exists, lies in the distant intellectual future" (Nagel 1974, 436). This is impossible to know, but it is a measure of the magnitude of the perceived challenge that it conjures up such pessimism among some of those who think deeply about the issues. It would

[12] There are a number of theoretical debates continuing in the philosophy of mind. Dating back to J.J.C. Smart's important (1959) paper "Sensations and Brain Processes," there is the view that seeks a theoretical reduction of paradigmatic mental states to brain states, just as we are able to reduce lightning to electrical discharges. More recently, Kari Theurer and John Bickle (2013) have revived something of a mechanistic approach to the reduction of the mental to the physical. Lockwood (1989) suggests that the existence of what we refer to as consciousness presents yet another challenge to the common-sense view of matter, just as does quantum theory.

be misleading, however, to suggest that such pessimism is the dominant view in the relevant literature.

To return to our theory of language, mainstream linguistic theory makes no mention of neurons. But it explains a vast range of phenomena more or less well, as the evidence is more or less compelling. That just happens to be the best anyone has been able to come up with, where it was possible to make progress. Where we do not seem to be able to make progress is by looking for inspiration in theory construction from what we know about brains. Of course, everyone would hope that one day, if we understand more and more about language and more and more about brains, we will be able to see how the two theories can constrain each other. This is perhaps where many scientists would situate their views. But if we take the further step of dictating how theory construction should proceed, then our notion of dualism is far from innocuous and it is also unwarranted. So far as we know, prior to Maxwell, no one berated those who tried to understand electricity for failing to frame their theories in terms of light or of magnetism, or vice-versa. The assumption that one side of a theoretical divide has to conceive itself in terms of the other side of the divide apparently applies uniquely to mind-brain relations; that is, it is a relic of a defunct version of dualism that serves no useful purpose at all (see Chomsky 1995 for discussion).

Versions of dualism persist, not only in discussions that reject linguistic theory as "brain averse" (Churchland's 2002 term), but also in discussions of 'eliminative materialism' and of qualia (Chalmers 1995), for example. We think that such views presuppose that there is some serviceable concept of 'physical' to which one can appeal. It would be interesting to know what characterization of 'physical' modern apologists for dualism have in mind. It would have to be something that can cover all objects that have extension, the fundamental forces (gravitational, electromagnetic and nuclear), electric and magnetic fields, quantum electrodynamics, and so on. Thus, perhaps all that can possibly be intended by the modern usage of 'dualism' is that we have our best theory of some aspect of mind on the one hand and our best theory of brain on the other and that they do not seem to share any properties. If so, then that is probably a fair characterization of where we are at this juncture in our intellectual history. But if all that is meant by 'physical' is the best that human minds have been able to come up with, that is, best theories, then no principled divide with the mental can be at issue, as we have argued; what is then at stake amounts to nothing more than a familiar, but very complex, theory-unification problem. Thus, we can point to our theory of brain and we can believe that the brain does the mind's work and wish that our theory of mind could match our theory of brain. But it doesn't just happen because we wish it to, and there is no point in belaboring a theory of some aspect of mind because it does not frame itself around what we know about brains.

In this chapter, we have provided a glimpse of the differing views of dualism from antiquity to the present. We cannot possibly do justice to the full complexity of the issues that any discussion of dualism gives rise to. Rather than attempt to go into every nook and cranny of 'if' and 'perhaps' and 'but', we have instead presented a particular view (one that readers might have encountered elsewhere, for example, in the writings by thinkers such as the mathematician, Morris Kline, or the linguist, Noam Chomsky); this view rejects dualism on the grounds that there is no sustainable mental-physical divide since we lack any useful concept of 'physical' and have not had one since Newton proposed action at a distance. Inquiry into mind-brain relations, on this view, is a problem of unifying theories, not a problem of reducing mind to matter, or any other formulation of the problem that seeks to salvage dualism. We hope that this at least provides a clear perspective that might generate worthwhile discussion.

References

Apostle H (1966) Aristotle's metaphysics, translated with commentaries. The Peripatetic Press, Grinnell

Apostle H (1969) Aristotle's physics, translated with commentaries. The Peripatetic Press, Grinnell

Apostle H, Gerson L (1991) Aristotle: selected works, 3rd edn. Peripatetic Press, Grinnell

Atherton M (1994) Women philosophers of the early modern period. Hackett Publishing, Indianapolis

Beretta A (2014) Joseph Priestley: an instructive 18th century perspective on the mind-body problem. In: Smith C, Whitaker H (eds) Brain and mind: essays in the history of neuroscience, Springer series in history, philosophy and theory of the life sciences. Springer, New York

Bright T (1586) *A treatise of melancholie*, containing the canvases therof, and reasons of the strange effects it worketh in our minds and bodies: with the phisicke cure, and spirituall consolation for such as haue thereto adioyned an afflicted conscience. Thomas Vautrollier, London

Chalmers D (1995) Facing up to the problem of consciousness. J Conscious Stud 2(3):200–219

Chalmers D (1996) The conscious mind. Oxford University Press, Oxford

Chomsky N (1995) Language and nature. Mind 104:1–61

Churchland PS (2002) Brain-wise: studies in neurophilosophy. MIT Press, Cambridge, MA

Cottingham J, Stoothoff R, Murdoch D (1984) The philosophical writings of Descartes vols I, II. Oxford University Press, Oxford.

Cottingham J, Stoothoff R, Murdoch D, Kenny A (1991) The philosophical writings of Descartes, vol 3. Oxford University Press, Oxford. Letter to Meyssonnier: (CSMK, p. 143); Letter to Mersenne: (CSMK, p. 145); Letter to Regius (CSMK, p. 181)

De la Mettrie JO (1748/1996) Machine man. In: Thomson A (ed) La Mettrie: machine man and other writings. Cambridge University Press, Cambridge, pp 1–39

Duerlinger J (2005) Plato's Sophist: a translation with a detailed account of its theses and arguments. Peter Lang Publishing, New York

Frankfort H, Frankfort HA (1949) The intellectual adventure of ancient man: an essay on speculative thought in the ancient near East. Penguin, Baltimore

Green CD, Groff PR (2003) Early psychological thought: ancient accounts of mind and soul. Praeger, Westport

Hamilton E, Cairns H (1961) The collected dialogues of Plato including the letters. Princeton University Press, Princeton

Jaynes J (1990) The origin of consciousness in the breakdown of the bicameral mind. Houghton Mifflin, Boston

Kirk GS, Raven JE (1971) The pre-socratic philosophers: a critical history with a selection of texts. Cambridge University Press, Cambridge

Kline M (1985) Mathematics and the search for knowledge. Oxford University Press, New York

Lewis CS (1964) The discarded image: an introduction to medieval and renaissance literature. Cambridge University Press, Cambridge

Locke J (1690/1975) Nidditch PH (ed) An essay concerning human understanding. Clarendon Press, Oxford
Lockwood M (1989) Mind, brain and the quantum: the compound "I". Basil Blackwell, Oxford
Margoliash D, Nusbaum HC (2009) Language: the perspective from organismal biology. Trends Cogn Sci 13:505–510
Matson W (1966) Why isn't the mind-body problem ancient? In: Feyerabend P, Maxwell G (eds) Mind, matter and method: essays in philosophy and science in honor of Herbert Feigl. University of Minnesota Press, Minneapolis
Nagel T (1974) What is it like to be a bat? Philos Rev 83:435–450
Pulvermüller F (2002) The neuroscience of language: on brain circuits of words and serial order. Cambridge University Press, New York
Russo J, Simon B (1968) Homeric psychology and the oral epic tradition. J Hist Ideas 29(4):483–498
Ryle G (2002) The concept of mind. University of Chicago Press, Chicago
Sabra AI (ed) (1983) The optics of Ibn al-Haytham, books I–II–III: on direct vision. The Arabic text, edited and with introduction, Arabic-Latin glossaries and concordance tables. National Council for Culture, Arts and Letters, Kuwait
Sabra AI (ed) (2002) The optics of Ibn al-Haytham. Edition of the Arabic text of books IV–V: on reflection and images seen by reflection. 2 vols. The National Council for Culture, Arts and Letters, Kuwait
Sandywell B (1996) The beginnings of European theorizing: logological investigations. Routledge, New York
Smart JJC (1959) Sensations and brain processes. Philos Rev 68:141–156
Theurer KL, Bickle J (2013) What's old is new again: Kemeny-Oppenheim reduction in current molecular neuroscience. Philos Sci 17(2):89–113
Vesalius A (1998–2009) On the fabric of the human body (trans: Richardson WF, Carman JB), 5 vols. Norman Publishing, San Francisco/Novato
Whitaker HA (2007) Was medieval cell doctrine more modern than we thought? In: Cohen H, Stemmer B (eds) Consciousness and cognition: fragments of the mind and brain. Elsevier Science, London, pp 45–51
White H (1963) An anatomy of kinship. Prentice-Hall, Englewood Cliffs