



From Consciousness to Brain-Sign: a Neurobiological Reconstruction

Philip Clapson¹

Received: 16 July 2020 / Accepted: 1 September 2020 / Published online: 25 September 2020
© The Author(s) 2020

Abstract

It may seem obvious we are conscious for we are certain we see, feel and think, but there is no accepted scientific account of these mental states as a brain condition. And since most neuroscientists assume consciousness and its supposed powers without explaining it, science is brought into question. That consciousness does not exist is here explained. The alternative, the theory of brain-sign, is outlined. It eliminates the quasi-divine knowledge properties of seeing, feeling and thinking. Brain-sign is a means/mechanism enabling collective action between organisms. Brain-sign signifies the shared world of that action. Signs are intrinsically physical and biologically ubiquitous. Brain-signs are derived moment-by-moment from the causal orientation of each brain towards others and the world. Interactive behaviour which is not predetermined (as in passing a cup of coffee) is characteristic of vertebrate species. Causality lies in the electrochemical operation of the brain. But identifying the changing world by brain-signs binds the causal states of those interacting into one unified operation. Brain-signing creatures, including humans, have no ‘sense’ they function this way. The world appears as seen. The ‘sense of seeing’, however, is the brain’s communicative activity in joint behaviour. Similarly for ‘feeling’. Language causality results from the transmission of compression waves or electromagnetic radiation from one brain to another altering the other’s causal orientation. The ‘sense of understanding’ words is the communicative state. The brain understands nothing, knows nothing, believes nothing. By replacing the prescientific notion of consciousness, brain-sign can enable a scientific path for brain science.

Keywords Brain-sign · Causal orientation · Consciousness · Interneural communication · Inter-organism communication · Neural interpretation

Introduction

A credibility gap lies at the heart of brain science. While consciousness supposedly illuminates the world, and language subjects its fabric to human mastery, it resists physical description. How it functions is ‘explained’ by substituting other words for it like ‘awareness’ or ‘experience’ which invoke similar non-physical concepts (e.g. Pennartz et al. 2019; Frith and Rees 2017). As a generality, neuroscientists identify active areas and conditions of the brain, then attribute to them mental categories as seeing, feeling or thinking, without specifying what they add to, or how they arise from, physical states.

At the same time, nonconscious neural activity is considered so extensive and productive of behaviour that what consciousness does beyond it is regularly asked with no agreed

account emerging (e.g. Anderson 2014; Oakley and Halligan 2017). However, it seems that describing brain function requires mentalist concepts and function (cognition, so-called) while mentality itself is unexplained.

A solution is offered by *brain-sign theory*. Brain-sign replaces consciousness as the brain phenomenon. It is a biophysical means of interneural communication by organisms about the world (including the organism itself) which facilitates collective action by signifying what in the world is jointly targeted. It results at each moment from the brain’s interpretation of its immediate *causal orientation* towards the world. Causal orientation both positions and can initiate bodily action. However, this is ‘invisible’ to us as brain-sign, for ‘we’ are wholly determined by the brain’s causal operation over which brain-sign has no control. Subjectivist mental concepts and lexicon, and the mind generally, are rejected as superfluous to science. Signs are biologically ubiquitous and intrinsically physical.

Since brain-sign is characterized by wholly physical processes, the mind-body problem is dissolved. A clear distinction is made between the phenomenon’s biophysical operation and its communicative content. Given the coherence of the

✉ Philip Clapson
p.clapson@bkk.ac.uk

¹ Department of Psychological Sciences, Birkbeck, University of London, London, UK

theory, proponents of consciousness needs demonstrate why brain-sign theory should not replace the intractable vagueness and implausibility of consciousness.

The three main criteria for developing the theory are to:

- 1 Provide foundations for brain science by replacing consciousness with brain-sign.
- 2 Distinguish the concepts and lexicon of brain science from that of (colloquial) mental states.
- 3 Improve ordinary discourse by a more lucid account of human being and other creatures.

After reviewing the status of consciousness and neuroscience ([A Background for Brain-Sign Theory](#)), this text outlines the theory ([The Theory of Brain-Sign](#)).

A Background for Brain-Sign Theory

Many problems are raised about consciousness: What is it? How can it be physical? How is it identified in the brain? How do neural states become knowledge? How do we deal with qualia? These questions assume consciousness exists. But consciousness has no scientific validation.

A comparison with the geocentric universe is instructive. On earth, the heavens appeared to move over us. This vastly exaggerated human significance and supported a belief in a divine realm above and a superhuman creator and ruler. Alternatives were anathema. Similarly, the history of mind and consciousness concerns a quasi-divine realm of knowledge by vision, feeling and the power of thought. For many, experience is supportive despite a literature which raises doubts. The following illustrates why a new approach is required.

Christof Koch says: ‘Many modern analytic philosophers of mind... find the existence of consciousness such an intolerable affront to what they believe should be a meaningless universe of matter and the void that they declare it to be an illusion.... If I have a tooth abscess, however, a sophisticated argument to persuade me that my pain is delusional will not lessen its torment one iota.... I have very little sympathy for this desperate solution to the mind-body problem’ (2018).

In their book *The Neurology of Consciousness*, Laureys et al. (2016) state that ‘Happily... this book clearly demonstrates that... consciousness [is] a viable subject for scientific study.’ (p. ix). But later this passage occurs in an article by Tononi et al. ‘Despite the wealth of evidence... it is difficult to converge on a circumscribed set of brain structures that are ‘minimally sufficient and jointly necessary’ for consciousness [quoting Crick and Koch 2003].... It is also important to keep in mind that, at this stage, we have no idea whether the elementary neural units that contribute to consciousness are local groups of neurons, such as cortical mini-columns, or individual neurons, and perhaps only neurons located in certain

layers or belonging to a particular class’ (p. 427). Towards the end is this. ‘We still need to understand why certain structures and processes have a privileged relationship with subjective experience’ (p. 445).

Nowhere is consciousness or subjective experience explained as either biologically necessary or physically plausible. The *Integrated Information Theory* (IIT) states that the essential properties are that it ‘exists intrinsically’ and is ‘extraordinarily informative’. It ‘exists for the experiencing subject rather than for an external observer’ (p. 445). No relation to physical or operational terms is offered.

Neuroscientist David Poeppel says: ‘The fact is that we have essentially no idea how the ‘stuff of thought’ relates to the ‘stuff of meat’, in the case of speech and language, and much the same is true in virtually all domains of higher cognition’ (2015, pp. 142–143). But Poeppel does not tell us what thought is.

In his 2014 book Michael Anderson says that ‘When neuroscientists start brandishing the ‘c’ word [consciousness], there are two predictable reactions: increased public interest and attention and increased scientific scrutiny and criticism’ (p. 109). The first involve ‘enthusiastic adherents’, the second ‘question whether we should continue wasting *our* energy figuring out why the brain appears to be wasting *its* energy’ (ibid.). However, Anderson offers no solution (personal communication); i.e. the problem is dismissed, not resolved, which is characteristic of cognitive neuroscience.

Stanislaus Dehaene reports that ‘When I was a student in the late 1980s, I was surprised to discover that during lab meetings, we were not allowed to use the C-word. We all studied consciousness in one way or another, of course, by asking human subjects to categorize what they had seen or to form mental images in darkness, but the word itself remained taboo’ (2014, p. 7).¹ Dehaene’s attempt to define it is this: ‘[1] The state of wakefulness, which varies when we fall asleep or wake up; [2] attention – the focussing of our mental resources onto a specific piece of information; and [3] conscious access – the fact that some of the attended information eventually enters our awareness and becomes reportable to others’ (ibid., p. 8). But this merely references other words for consciousness: ‘wakefulness’, ‘mental resources’, ‘attended information’, ‘awareness’, ‘conscious access’. It does not explain in a more primary and biologically illuminating way.

Lancelot Law Whyte wrote these prescient words in his 1960 book: ‘What does mental mean?... It is here the problem lies. *No one yet knows how properly to define ‘mental,’* perhaps because this can only be done within a valid monism’ (p. 63).

Whyte’s words illustrate that (1) the problem does not begin with identifying consciousness in the brain for it has no

¹ It is still normal for manuals on neuroscience or neurobiology to make no reference to consciousness but to assume that physical states and processes nonetheless result in (unexplained) experience or sensation (e.g. Luo 2016).

scientific definition. (2) There is no adequate explanation of what a definition would be. (3) Nor what language *is* that could establish it.

Given this context, the following maps the struggle to address the topic.

The Perfidious Brain

While proponents of consciousness assume that experience is self-revelatory they talk of the difficulty of defining it. Dehaene, for example, says ‘The word *consciousness*...is loaded with fuzzy meanings covering a broad range of complex phenomena’ (2014, p. 8). But a science of consciousness would require a clear ontological status and biological function *as* the brain. If the brain phenomenon is not consciousness then defining it would not involve circularity because the knowledge, which consciousness hypothetically provides, would not exist.

A response might be: ‘But where would our knowledge come from? I know the world is right there before me.’ Which echoes Koch’s complaint: ‘A sophisticated argument to persuade me that my pain is delusional will not lessen its torment one iota.’ So the first stage entails suppressing the conviction that ‘the world is right there before me,’ and accepting that ‘my torment’ must be explicable in scientific terms. Otherwise the organism is elevated from biology by unknown properties. *Certainty* about my pain and vision must also be explained.²

Of course, if humans are not so elevated, they are not as generally supposed. Herein lies difficulty. For the freedom effected by consciousness (‘I make decisions’) is an escape from biophysical determinism.³ How could our behaviour be determined yet our experience tells us otherwise? So strong is the grip of consciousness, the brain is accused of deceiving us. A few examples.

Not Getting to the Matter

In his book *The User Illusion*, subtitled *Cutting Consciousness Down to Size*, Tor Nørretranders intends to demonstrate what consciousness actually does. ‘Consciousness is depth but is experienced as surface’ ((1991), 1998, p. 288). ‘We experience

² Overtuning Galen Strawson, ‘There occurred in the twentieth century the most remarkable episode in the whole history of ideas – the whole history of human thought. A number of thinkers denied the existence of something we know with certainty to exist: consciousness, conscious experience.’ Isaiah Berlin Lectures, Oxford University (2017).

³ e.g. Antonio Damasio: ‘Consciousness is just the latest and most sophisticated means of generating adequate responses to an environment...by making way for the creation of novel responses in the sort of environment which an organism has not been designed to match, in terms of automated responses’ (1999, p. 304) (slightly reconstructed) which it does by ‘the world of planning, the world of formulation of scenarios and prediction of outcomes’ (p. 303).

not the raw sensory data but a simulation of the sensation.... We do not experience things themselves. We sense them. We do not experience the sensation. We experience the simulation of the sensation...an illusion’ (p. 289). The word ‘illusion’ applies because we take it we experience what is so between ourselves and the world, whereas the brain has made a construct both by working over masses of sensory data, and discarding the irrelevant. We are presented with a comprehensible result. So the argument goes.

But the language is crucially muddled. ‘Raw sensory data’ is related to sensory organs, but a sensation occurs as an experiencer’s experience by definition. What Nørretranders means to say is that sensation is not of the raw sensory data. Instead he first says that ‘we do not experience the sensation’, which (assuming consciousness) we do (‘we sense them’ he says), but then he says ‘we experience the simulation of the sensation’ which is nonsense. The appropriate question is that given sensory data (more precisely, sensory events) are legitimate expressions about sensory interfaces with the physical world, what, as physicality, is the experiencer’s experience of sensation? i.e. what is sensation feeling supposed to be? For without explanation, its elements (experience, sensation, feeling) remain undefined—along with their host, consciousness. And why is it an illusion—to what, to whom? Nørretranders does not reach the heart of the matter.

Not Getting to Biology

In a footnote in his 2017 book *From Bacteria to Bach and Back* (p. 335), Daniel Dennett refers to Nørretranders’ book as published in the same year as his *Consciousness Explained* (1991), so there was no mutual reference. In Chapter 14 of Dennett’s 2017 book, titled ‘Consciousness as an Evolved User-Illusion’, he expounds on his computer-model version of consciousness.

‘We can list the properties of the tokens on the computer desk-top: blue rectangular ‘files’; a black arrow-shaped cursor; a yellow highlighted word in black Times Roman 12-point font.... What are the corresponding properties of these internal, re-identifiable private tokens in our brains? We don’t know – yet.... Close your eyes and imagine a blue capital A.... You just created a token in your brain, but we can be sure it isn’t blue, any more than the tokens of ‘o’ that occur in a word-processing file are round. The tokenings occur in the activity of neural circuits, and they have an important role to play in directing attention, arousing associated tokens, and modulating many cognitive activities’ (p. 347).

The term ‘user illusion’ comes from the computer industry. Computer and brain functioning are made analogous. The ‘user’ (the human being) need not be acquainted with the work

the computer does to facilitate ease of (e.g.) screen interface. Likewise, consciousness is, as it were, the brain's interface to itself, the intelligible state of the brain. We can tell that because, sourced by the neural brain, the world is intelligible to us as a mental construct.

Dennett explains consciousness by proposing it emerges from an accumulating hierarchy of simpler physical states, gradually increasing the range of its competence.⁴ (He referred to it in 1991 as a 'virtual machine', and still does in 2017.) The problem, however, is not, as proponents of consciousness say, that Dennett has not explained consciousness because it has some kind of being which amassing operational circuits in a hierarchy cannot just become (e.g. David Chalmers 1996; Ned Block, see below). It is that the brain phenomenon (aka consciousness) has not been given a biofunctional role. Dennett wants to demystify consciousness. But the computer analogy is not explanatory. As with Nørretranders, what is consciousness supposed to do that the brain *qua* brain states could not? The answer requires a biological explanation, not a computer analogy (cf. Chomsky 2016, p. 29).

The Explanatory Block of Mentalism

In his 2017 book, Dennett addresses the work of Daniel Wegner and his *The Illusion of Conscious Will* (2002). Here is a relevant passage from that 'groundbreaking' book. 'We can never be sure that our thoughts cause our actions, as there could always be causes of which we were unaware that have produced both the thoughts and the actions.... As Nisbett and Wilson (1977) have observed, the occurrence of a mental process does not guarantee the individual any special knowledge of the mechanism of this process. Instead, the person seeking self-insight must employ a priori causal theories to account for his or her own psychological operations.... Conscious will is not a direct perception of the relation but rather a feeling based on the causal inference one makes about the data that do become available to consciousness – the thought and the observed act' (pp. 66–67).

As with Nørretranders, we are in a terminological black hole. First of all, consciousness exists but without definition. Then, it is given attributes: 'a mental process', 'knowledge', 'self-insight' (i.e. introspection), the (necessarily conscious) 'person' who has 'a priori causal theories' and 'psychological operations', together with 'a feeling based on causal inference' that 'one [i.e. the mental subject] makes' 'about the data which become available to consciousness' with a 'thought' and an 'observed act'.

⁴ As Dennett says in his 1991 book, 'The way to discharge an intelligence that is too big for our theory is to replace it with an ultimately mechanical fabric of semi-independent semi-intelligences acting in concert' (p. 251).

Wegner's aim is to escape the involvement of mental life in causal activity because causality belongs to the physical world: 'the [physical] mechanism of the process'. But the explanatory text wholly depends upon mental life, else how can the nature of human being be described? But mental life, specifically consciousness, is not addressed. The index entry for 'conscious mind' sends us to this passage. 'The definition of will as an experience means that we are very likely to appreciate conscious will in ourselves because we are, of course, privy to our own experiences and are happy to yap about them all day' (p. 11). The 'of course' indicates that little is revealed as science. Being 'privy to our own experiences' is opaque. What is identified by 'our'? Is it a sense of self, and what is that vis-à-vis the brain? Then there is 'own'—that which belongs (presumably) to me. But what is myself: how is it owned? How are we 'privy'? How is privy-ness made available? What is an 'experience'?... The book concerns the illusion of conscious will, but the foundational element, consciousness (of conscious will), is never explained. It is assumed in the history of the word and concept only describable by other terms of itself.

Could Consciousness Play Tricks?

In an attempt to deal with consciousness and the notion of *qualia*, or qualitative states, Dennett uses David Hume's seminal comments on causation. 'We seem to *see* and *hear* and *feel* causation every day, Hume notes' (2017, p. 354). But Hume said that this is a special case of the mind's "great propensity to spread itself on external objects (1739, I:xiv)".... It survives to this day in the typically unexamined assumption that all perceptual representations must be flowing inbound from outside' (ibid., p. 355). But 'you can't find intrinsic sweetness by examining the molecular structure of glucose: look instead to the details in the brain of sweetness seekers' (pp. 355–356). This is the brain's 'benign illusion'. And our brains 'have tricked us into having the conviction... that there seems to be an intrinsically wonderful but otherwise indescribable property [i.e. *qualia*] in some edible things: sweetness' (ibid., p. 356).⁵

The phrase 'our brains trick us' is now a common expression. Nick Chater, who claims there is no unconscious mind, reprises the arguments. 'So, except in a rather uninteresting sense, we aren't really conscious of numbers, apples, people, or anything else – we're conscious of our interpretations of sensory experience (including inner speech) and nothing more. In this light, the tower of levels of consciousness, each built on the last, collapses. It is one more trick played on us by the brain' (2018, p. 185).

⁵ Kant incorporated the notion of causality into the mind's Transcendental Categories of Understanding in the *First Critique* (1781, 1787), responding to Hume's critique.

But is it likely the brain, as consciousness, plays tricks on us? (Which ‘us’?) What possible survival benefit could that entail?

Current Non-scientific Nomenclature

Developing a scientific theory of the brain phenomenon is of a different order from a theory of the physical universe (heliocentrism, molecular chemistry, quantum field theory). Although humans and other creatures may be purely physical, theories about ourselves are generated by our brain’s capacity for theorizing, and it is that which requires scrupulous exploration.

Explanation Which Does Not Explain

In an article titled ‘Consciousness and Conceptual Clarity’, Ned Block says that: ‘The lesson to be drawn is that isolating consciousness in the brain may depend more on being clear about what we are looking for than on massive investments in new technology’ (2015, p. 175). He follows with a paraphrase of Kant. ‘Concepts without data are empty; data without concepts are blind’ (ibid.). This is a seminal point. Does Block achieve it?

He illustrates his point about concept making by contrasting Dehaene’s theory of global broadcasting with the results of experimental data from Wolfgang Einhäuser’s lab (Frässle et al. 2014).

Dehaene, Jean-Pierre Changeux and colleagues (2011) advanced a global neuronal workspace theory of consciousness,⁶ in which neural coalitions in the rear of the brain (the primary visual cortices) compete for dominance, the winners linking long range with the prefrontal cortex where cognitive functions are located, generating feedback to the visual cortices and thence ignition of the workspace. The workspace becomes widely available to other brain processes. For Dehaene et al. the workspace is consciousness, i.e. it is cognitive functioning. However, the Einhäuser’s lab experiments on binocular rivalry, rather than depending on the subject’s account of their (cognitive) experience, introduced eye-tracking technology which identified eye movements (left or right). A response resulted from subjects pressing a button without conscious cognition having taken place because they were not required to report it. It was therefore not associated with the global broadcasting theory; frontal processing had not occurred (Block 2015, pp. 174–175). Thus phenomenal consciousness (Block 1995), i.e. conscious activity, was apparently identified without cognition. Block termed cognitive activity access consciousness. As Block says in the (2015) article, ‘Lamme (2003), Zeki, and I do not think that phenomenal consciousness has

no information processing role. We think that consciousness greases the wheels of cognition but can obtain without it’ (p. 167). In other words, access consciousness (pressing the button from eye-tracking) is not conscious, but is so when phenomenal consciousness is also active. What ‘greasing the wheels’ entails is not discussed.

These theoreticians have different accounts about a supposedly fundamental biological construct. But neither tells us what being conscious does, i.e. why it is necessary beyond the causal power of physical brain states. We are not even told by Block what consciousness is, ontologically. He quotes Dehaene concerning his (Block’s) position. ‘The hypothetical concept of qualia...will be viewed as a peculiar idea of the prescientific era’ (2014, p. 221). But since Dehaene does not tell us why ‘we’ as consciousness are an ignition of the brain rather than just physical processes operating, there is no guide for the matter.⁷

Is the Brain Phenomenon Really Consciousness?

There is another dimension. Why does consciousness exist at all? Before Freud, Nietzsche stated that ‘For the longest time, conscious thought was considered thought itself. Only now does the truth dawn on us that by far the greatest part of our spirit’s activity remains unconscious and unfelt’ (1974, p. 262). Under Freud’s aegis, this generated a profound change in human self-conception in the twentieth century. The causal soul was replaced by anatomical function, though Freud failed to establish how. Nietzsche had an additional insight (cf. Hume/Kant). ‘Explanation’ is what we call it, but it is ‘description’ that distinguishes us from older stages of knowledge and science... How could we explain anything? We operate only with things that do not exist’ (ibid., p. 172)—i.e. mental constructs. This is what Nørretranders et al. aim at: the illusory nature of consciousness. (Use of ‘explain’ in this text assumes Nietzsche’s point.)

However, John Bargh said in 2005, ‘If we are capable of doing something effectively through nonconscious means, that something would likely not be the primary function for which we evolved consciousness’ (p. 52). But Bargh hypothesises that ‘metacognitive consciousness [i.e. being aware of ‘my’ conscious content and therefore being able to influence it] is the workplace where one can assemble and combine the various components of complex-motor skills. This is a development of the human species because [quoting Donald 2001, p. 8] ‘whereas most other species depend upon their built-in demons to do their mental work for them, we can build our own demons’ (p. 53). (The tenor of the Baars/Dehaene workspaces.) Bargh continues: ‘The purpose of

⁶ Derived from the global workspace theory of consciousness developed by Bernard Baars (e.g. 1997), as acknowledged in the text.

⁷ Dehaene et al. retain this position in 2017. ‘What we call ‘consciousness’ results from specific types of information-processing computations, physically realized by the hardware of the brain.’ Also introduced is self-monitoring.

consciousness – why it evolved – may be for the assemblage of complex non-conscious skills... People have the capacity of building ever more automatic ‘demons’ that fit their own environment, needs and purposes. As William James (1890) argued, consciousness drops out of the processes where it is no longer needed’ (ibid.).⁸

This has ironic consequences for psychology, for ‘the evolved purpose of consciousness turns out to be the creation of ever more complex nonconscious processes’ since, according to Bargh, consciousness has a ‘limited capacity nature’ (referencing Baumeister et al. 1998), (ibid.).

But that implies psychology or cognitive science cannot be neuroscience’s sought-for descriptive account of brain function, for psychology depends upon the causal properties of mental states—seeing, feeling, thinking, willing, motivation. Indeed, Peter Carruthers, in his Interpretive Sensory Access Theory (ISA), and from extensive analysis, says: ‘If there are no conscious decisions then...there is no...conscious agency.... If there is no conscious agency, then there are no consciousness agents’ (2011, p. 379). But he does not then expunge consciousness. He avoids commitments regarding its nature (ibid., p. 373).

None of this solves the underlying question, why what has been supposed as consciousness need exist. But authors do not break free from it. The question remains: Why does all brain work not occur without consciousness, particularly when it is apparently so limited?

Does Consciousness Make Sense?

Moving forward, David Melnikoff with John Bargh in 2018, debunk the conscious/unconscious typology of mental processes. Type 1 is efficient, unintentional, uncontrollable and unconscious. Type 2 is inefficient, intentional, controllable and conscious. They say ‘This...typology has grown more popular with each passing decade. In just the past 5 years it has shaped empirical and theoretical work’ on numerous areas of research, e.g. emotion, religiosity, interview bias, judgement and decision making. They continue: ‘Popularity of this magnitude is typically reserved for ideas that have withstood decades of conceptual scrutiny and empirical vetting, so it is no surprise that the Type 1/Type 2 distinction has a reputation among many researchers of being uncontroversial, even axiomatic. But this reputation, it turns out, is undeserved.’ On the contrary ‘There is no evidence that processing features cluster together into two groups; and there is substantial evidence they do not.’ Each of the areas is explored in the light of experimental research. They say, with some urgency, ‘It is time that we...come to terms with these issues. With organisations like the World Bank and Institute of Medicine now

endorsing our highly speculative and frequently misleading typology, we cannot afford to wait.’

They point out that with four sets of binary features in each of the two types, there are 2⁴ possible combinations. However, the typology proposes only two, leaving 14 non-applicable. Unlikely, they claim. Intentional goes with conscious and controllable presumably on the assumption that consciousness has the power of intending and controlling what is done. Unintentional, unconscious and uncontrollable are the inverse.

Here is one of their examples. ‘A bat and a ball cost \$1.10 in total. The bat costs \$1 more than the ball. How much does the ball cost?’ People almost invariably generate an initial answer of 10 cents, but the correct answer is 5 cents.’ The immediacy of the response is uncontrolled (Type 1) while at the same time is consciously intended (Type 2), thus obviously breaking the typology. It is not that the wrong answer is caused by the failure to work it out correctly but rather that the conscious intention is to answer but is uncontrolled (unconscious).

One might say (which the authors do not) that the wrong answer is generated unconsciously (Type 1), which is why it is uncontrolled and, as it were, passes seamlessly through consciousness (Type 2) without further analysis (i.e. it is not controlled). For the immediate answer to the question is ‘obvious’ from the question’s phrasing (\$1.10 cost and \$1 more-than, leaving 10 cents). Further thought is required to get the right answer.

But is thinking conscious, or is it, as Bargh hypothesised in 2005, a more elaborate unconscious activity pre-generated via consciousness? If so, here is an alternative analysis. While a correct answer might seem to result from consciousness (Type 2), since it evidently passes through it, it is the unconscious that can generate a delayed but ultimately efficient correct answer (Type 1), which would align with the typology if it too was uncontrolled (Type 1). That is, consciousness has no immediate impact at all despite the experimental assumption it is intended. Thus the typology fails, as the authors are aware but do not analyse in this example.

They summarise: ‘Given that there are at least three dissociable ways in which a process can be unconscious, it makes little conceptual sense to talk about consciousness as a unitary processing feature that can co-occur with other features.’ This again undermines the disciplines of psychology and cognitive science, for what is demonstrated is the total absence of a viable mentalist vocabulary (conscious/unconscious) related to scientific concepts.

Summary

Mental life is opaque (Carruthers’ word) and problematic rather than precise and relevant to neuroscience. It lacks scientific definitions. Its function has no agreed interpretation. It is divisive since either our very being appears to demand it (Koch,

⁸ These ideas already exist in Nietzsche’s writings.

Strawson), or scientific impetus forces implausible solutions (Dennett, Dehaene, Tononi). Indeed, Dehaene says that ‘no experiment will...show how the hundred billion neurons in the human brain fire at the moment of conscious perception. Only mathematical theory can explain how the mental reduces to the neural’ (2014, pp. 162–163). In a similar vein David Poeppel says: ‘Bridging virtually all domains of higher cognition and neurobiology in an explanatory fashion requires the formulation of computationally explicit linking hypotheses’ (2015, pp. 143, slightly rearranged). Mathematics is a vital tool. However, it is the mentalist assumption that is the source of the crisis in theorising, not the absence of computation. As Bargh and Melnikoff say, the current conscious/unconscious schemata is ‘systematically thwarting scientific progress.’ The following is a brief outline of an alternative.

The Theory of Brain-Sign

Science requires an account of the brain phenomenon with a plausible physical schema. While mind, soul and spirit address the mystery of human being, they are linked to notions of divinity; in the case of mind, what Paul Churchland dubbed ‘folk psychology’ (1981).

Here is the key analytical issue, referenced in ‘[Current Non-scientific Nomenclature](#)’. We are the object of scientific enquiry (mind/consciousness) and the means of enquiry (as mind/consciousness). Consciousness, as means and object, is presupposed without identifying whether consciousness is actually consciousness (‘[Is the Brain Phenomenon Really Consciousness?](#)’ and ‘[Does Consciousness Make Sense?](#)’). Is the ‘we’ that enquires, apparent in Koch’s cry, a scientific entity? ‘A sophisticated argument to persuade me that my pain is delusional will not lessen its torment one iota.’ Koch does not justify the ‘me’ scientifically. It is a subject harbouring the divinity (the knower knowing) within the mentalist lexicon.

Actually, there are three topics to be dealt with, not two. They are (1) the function of our ‘presence’ in the universe, i.e. what are ‘we’ as the brain phenomenon? (2) The physical nature of that phenomenon fulfilling a coherent biophysical function, and therefore (3) the physical actuality of the brain phenomenon.

What is required is a way of reconstructing concepts and evidence so that a plausible scientific model appears from the ruins of a prescientific construct. Of course, given the history, this will require a fundamental change in our self-conception.

Locating Brain-Sign

Currently, and for thousands of years, humans have taken communication with others for granted because the world is immediately available as seen. Given this is impossible, for the brain is isolated in each organism’s skull denying it direct

access to the world, consciousness, as a theory, provides the resolution by positing seeing and hearing (or the sense of seeing and hearing, cf. Nørretranders). But if consciousness is a myth, how could brains communicate about the world?

The answer is by signification, for signs are inherently physical. Consider the death’s-head hawkmoth. The European variety can move about devouring honey in the hive of honeybees because it emits chemicals (a pheromone) whose properties are akin to those of bees, and it is thus ‘unnoticed’ (Moritz et al. 1991). The sign, the chemistry, does not facilitate communal action but it does render moths insensible to bees. More relevantly, a tiny brown pufferfish, identified off the coast of Japan, creates in 7 days elaborate symmetrical wheel-shaped designs in the sand with its fins to attract a female to lay eggs in the middle which it then fertilizes (Barrington et al. 2014, pp. 184–187). The sign engenders collective action which aids the species survival. How do these creatures know to do this? They do not. Their actions are generated by Donald’s ‘built-in demons’ developed in evolutionary time. Dennett (2017) refers to this activity as competence, not comprehension.

But there is crucial differentiation in organisms’ biophysical sign usage. Bargh/Donald state it is between built-in demons and humans who ‘can build [their] own demons’. The use of pheromone chemistry by insects (and others) to generate cooperative action, or to defend or deceive, is well-documented. This sign communicates at a comparatively primitive biological level, as well as for humans. But it is functionally constrained. More elaborately, honeybees convey to others the route to pollen by the angular relation to the sun (and other factors) in their ‘dance’ signs. This is accomplished by altering receivers’ brain states via intervening electromagnetic radiation and compression waves. But whilst more elaborate than pheromone release, the function remains precise. The dance is an instruction. The pufferfish employs further complexity, for whilst the wheel design is an instruction if followed, it appears to have an ‘acceptance-or-not’ role for the female. Moreover, the sandscape is modified to achieve a result, which is a remarkable technical development.

An evolutionary transition is taking place. Organism cooperation moves from instruction (molecular or behavioural) to an operation which is neither genetically preestablished or mimicked. It demands continual adaptive behaviour, as in one person passing a cup of coffee to another, i.e. without predetermined responses. Dynamic interaction. Not only is this more elaborate, it needs a means/mechanism to facilitate it. That is brain-sign.

The Missing Link—Causal Orientation

The brain phenomenon, brain-sign, represents the world and the response of the organism to the world; what mentalism supposes as seeing and emotion. But (mental) seeing and

emotion are held to spring from the physical brain as different kinds of entities from the brain and each other. Mentalism is prolific, Dehaene's 'fuzzy meanings covering a broad range of complex phenomena.' Seeing, hearing, sensing, smelling, thinking, judging, discriminating, pain, depression, excitement, boredom, passion, shame, love, contempt, guilt, hope, confidence, etc. These are understood, not merely as personal attitudes, but, hypothetically, states of the mind/brain.

Brain-sign theory, however, has a straightforward explanation for the origin of them all. For something determines what segment of the world (including the body) is to be represented. Assuming the brain itself causes the organism to act, it is the brain's immediate causal orientation towards the world (i.e. the brain's activated structures and electro-chemical status). In the case of seeing, what causes what appears are the specific elements of the world of the current causal orientation, including the organism itself. (Which is why the world can be seen in different ways depending on the brain's causal orientation.) But the brain does not see because its causal orientation has already determined what will happen. The brain phenomenon is the brain's interpretation of its causal orientation towards the world as that world. (Which is supported by the experiments of Benjamin Libet (e.g. 1983) in which a subject's 'chosen' action neurally precedes awareness of it.) But why would it do that if the image is not involved in its action?

Brain-sign is a means/mechanism by which brains jointly establish the domain of collective action from the successive causal orientations. If A passes B a cup of coffee, each brain continually interprets its shifting causal orientation so signifying the cup and the ongoing interaction within the local environment. The cup is identified between hands and bodies from the time of inception to receipt. Thus brain-sign is not an external instruction one to another, as the bee dance or pufferfish wheel. It is the content of the in-the-world transaction which each causal brain must continually establish to effect it. This includes, of course, the 'sense' of each individual being in that world.

Still, the question might be asked: If causality belongs in the physical brain operation, why do brains have to signify the world in which that operation takes place? The answer is that both the complexity and precision of interaction demands sustaining what the transaction is about. Brain-sign is the mutual reference for the brain/body action, though it is obviously not the actual world (cf. Kant, Nørretranders, Dennett, et al.). In action, they are one biophysical unit, not two or more.

Foundation and Evolution

Brain-sign did not develop so that complex interactions could take place. That would counter evolutionary theory. The hypothesis is that the internally generated representation of the world and responses to it result from the increasing causal possibilities of evolving brains in relation to the world,

particularly in vertebrate species. In numerous situations, representations became sustained as the world-object of causal brain structures—and thence, under natural selection, come to act as a communication medium/mechanism, greatly expanding possible behavioural complexity. Representations do not add to causality for the organism itself: they become the communicative means of the interaction process.⁹ For the pufferfish, the wheel pattern is communication in the mating act. It operates in the time-sequence of mating. However, increasing levels of cooperative interaction in mating behaviour occur. Many species of birds, for example, build nests for egg depository which is interactive, though the female is often dominant. In evolutionary time interaction occurs with much more advanced levels of reciprocation—for elephants, dolphins, apes and humans. Thus Bargh's underminings of consciousness are given a scientific reconstruction, and science itself a plausible physical genesis (i.e. with no conscious/unconscious dichotomy).

This is a new domain of biological explanation. Credibility may be enhanced by a biological parallel between brain-sign as representational states of world objects, and the cellular 'colouration' and design of the chameleon and cuttlefish, or the surface markings of butterflies acting as deception. Other body structures, too, in which aspects of the world are 'mimicked' for defence, for example, stick and leaf insects, the *Phasmatodea* (Butler 2012). Brain structures are certainly more elaborate than body surface structures, not least because they continually change, as does causal orientation. How the brain mechanism is constructed is as yet unspecified. Tononi et al. say this about consciousness (above); but they have the more intractable problem of explaining consciousness itself.

Coincident to the neural representations of the world, brain responses to the world are internalised for signification (so-called emotion, etc.). These are temporally linked to world representations but are also sourced from the organism's causal orientation.

Brain-Sign: an Outline

Causal orientation is the crucial link from brain to brain-sign. Organisms do not act from seeing, thinking, feeling or belief. They have no motives for what they do. These explanations are brain generated without scientific constraint: they are pre-scientific. Brain-sign theory initiates a new biophysical account of organisms. Here are fundamental elements of brain-sign structure.

⁹ By contrast with, e.g. Tomasello 2014; see Clapson 2016. Also this discounts references Dennett makes (2017, p. 342) to McFarland (1989), Dawkins and Krebs (1978) and others, in which the ability not to reveal the organism's 'current state to competitive organisms' is as significant as the ability to communicate. For generating representations is not required for that behavioural operation, nor is it explained how representations could be causal, i.e. the mind-body problem.

- i Categories-of-the-world
- j The biophysical marker
- k Categories-of-interaction
- l Brain-sign language

Perception Redesigned

From infancy, the organism explores the world. But the infant does not learn to see. The brain/body becomes acquainted with objects by establishing, in neural structures (termed assemblies or networks), causal responses to surfaces, substantiality, construction, behaviour characteristics, native environment and use. By sensory input and interaction with the world (eyes, ears, bodily contact, tongue), and thence brain association, neural instruction about action in the world is built. As gradually developed, the brain interprets its causal orientation towards the world thus generating brain-signs—portrayals of the world and the organism’s response to it, which signify in joint behaviour. Parents and children begin to engage beyond purely autonomic activity.

As a theory, mentalism’s seeming to see is replaced by an ‘invisible’ capacity for interneural communicative acquaintance with the world implicit in everyday behaviour. It is not apparent to our lived lives. Thus the account moves from conscious/unconscious duality to causal orientation and interneural communication—the sought-for scientific monism (ref. Lancelot Law Whyte). Brain-signs are part of the organism’s survival arsenal.

The sense of seeing (and other senses) is the signification of the vast assembly of neural structures that enable behavioural and cooperative possibilities of the brain and body.¹⁰ It is a communicative shorthand for the brain’s operational capacity in numerous potential situations. What is predominant in ‘seeing’ is the image. But the organism’s so-called sense of seeing is actually the brain’s outward act of neural communication, whether another organism is present or not. The ‘sense of seeing’ an apple conveys causal possibilities in relation to apples as physicality: eatability, grown on trees, makes cider, what Eve passed to Adam, and so on, the particularities dependent on the brain’s immediate causal orientation. So, a signified wall entails the behavioural feature of not walking into it. The brain will not be communicating actually if no other organism is present, but its biological status is communicative. Thus the mentalist terms ‘consciousness’,

¹⁰ Edmund Husserl developed his notion of the life-world (*Lebenswelt*) in which we exist in the world with its objects around us—the experience we take as normal and simply assume. His last published book, *Crisis* ((1954), 1989), influenced by Martin Heidegger’s *Being and Time* ((1927), 1962), distinguishes the experienced world from the scientific account of the world. Phenomenology sets out to describe the experienced characteristics. Brain-sign, as the portrayal of the world from the brain’s causal orientation, is scientifically founded and eschews phenomenology as scientifically incoherent.

‘experience’, ‘awareness’ etc. have no underlying neural associations. Brain-sign theory states that there is nothing in the world the brain can know (cf. Nietzsche et al.).

A luminous example of brain-sign’s improvement over psychology occurs with the notion of attention. In the well-known experiments of Daniel Simons and Christopher Chabris (1999, 2010), individuals concentrating on a video game of basketball fail to notice a person in a gorilla suit appearing, chest pounding, and departing. The phenomenon was given the term ‘inattention blindness’. It even occurs if viewers are warned of an irrelevant event beforehand (Chabris and Simons 2010). Brain-sign theory rejects inattention blindness: the brain sees nothing so it cannot be blind. The brain’s causal orientation is the basketball game which the resulting brain-sign signifies. This straightforward neuroscientific explanation replaces what has seemed a mental fallibility. Indeed, the flyer for the Chabris/Simon 2010 book states that ‘Our minds don’t work the way we think they do.’ Yet the mind is retained as a valid entity—ironically personifying the flyer. Similarly, Carruthers refers to this experiment in an interview (2018) but, despite asserting that ‘consciousness is not what we generally think it is,’ he maintains it, so fixated is human culture.

So returning to ‘*The Theory of Brain-Sign*’, we can say that (1) ‘Our presence in the universe’ is a biophysical condition of neural communication; (2) ‘The physical ontology of the phenomenon’ is that of a sign; and (3) ‘The physical actuality of the brain phenomenon’ is structures and conditions of the brain, as yet to be determined. Of course, brain-signs between organisms are not identical. They are adequate for a bio-communicative role.

‘Perceptions’ in their many forms (sight, hearing, touch, taste) are termed, as brain-sign content, categories-of-the-world for obvious reasons. They are not input to cognitive mechanics (Kant’s engine of reason) or any form of ‘what it’s like’ (Nagel 1974). They are output serving communication about the world of causal orientation. The organism does not perceive: the world comes into being as brain-sign for interneural communication.

*

It is appropriate to place the brain-sign account within the literature. (If the reader wishes to continue the brain-sign narrative, skip to the next section.) Firstly, the use of the term ‘seeing’ by contrast with ‘perception’. Whilst they are used interchangeably, perception is often associated with a more complex condition. Jan Koenderink and Joachim Krueger, in their 2017 article, express this difference. The word ‘seeing’, or ‘vision’ as the authors say, assumes the world is directly and accurately present to the observer. ‘The central problem [for rationality] is the present emphasis on inverse optics – the [supposedly] objective nature of objects and environments [as seen].’ By contrast, they propose the world is a construct in which the organism’s species, and thus sensory modalities, are incorporated together with its particular history (‘expectations,

conjectures, and theories’) which affect the perceiving state (cf. Nørretranders, et al.).

Brain-sign theory certainly agrees the brain phenomenon is a construct: it derives from the causal orientation of the organism’s brain. But whilst distancing their account from rationality’s non-biological ‘all-seeing eye’, they preserve mentalism and its elevation from the physical. No explanation is offered for its causality.

The second and linked topic is enaction or enactivism. As John Stewart puts it, “‘The world’ as it can be diversely known by living organisms from bacteria to contemporary humans is actually brought about, ‘enacted’, by the cognitive organism itself” (2014, p. 27). This is the so-called embodied mind (or cognition, the topic presaged in the book by Varela et al. 1991). Alva Noë, a pioneer of the approach, says that ‘At the ground of our encounter with...different objects – appearances in one modality or another – is sensorimotor skill’ (2004 p. 107). Being embodied means that mental states are not disembodied entities (Descartes) or with irreducible qualitative properties (Chalmers). Appearances result directly from bodily actions in the world.

Brain-sign theory sympathises. There are no causal mental states and brain-signs are generated from brain/body causality. Moreover, Stewart comments that ‘the majority of epistemological positions...share a commitment to objectivism’ and this makes enactivism ‘an unusual point of view’. However, neither Noë nor Stewart explain the biophysical *function* of world appearances. Why are they there at all? Stewart even says, ‘We humans are profoundly social beings’ (ibid.) without arriving at brain-sign theory. This characterises a diverse literature,¹¹ which probably accounts for its failure to gain unified assent—Thomas Kuhn’s ‘normal science’ (1962).

This failure is mentioned by Matthew Cobb in his 2020 book (p. 251) in relation to the book by neuroscientist György Buzsáki (2019). Buzsáki, working in the action-based genre, says ‘Cognition can be understood only as a social phenomenon that transcends the brain of an individual’ (p. 228). But his fascinating account is still consciousness-bound, not brain-sign.

The Subject (I, Self) Redesigned

For mentalism, the I is crucial for experience, and is conceptually central; hence Koch’s ‘my pain is [*not*] delusional’. But Koch’s discomfort is a communicative neural event. The causal orientation from which ‘pain’ is interpreted signifies the organism should do something about it. This was Descartes’ explanation, but he proposed it was divinely sourced (effectively endorsed by Chalmers’ ‘hard problem’ and Block’s phenomenal consciousness). ‘Pain’ per se is a category-of-

the-world. The response to pain, the ‘discomfort’, will be addressed in the next section.

In brain-sign communication, each organism is identified as unique. As content, ‘pain’ is conjoined with the biological identification of that organism, the biophysical marker, which replaces the mental subject. Marker and ‘pain’ exist together as biophysical communication. Koch’s ‘sense of himself experiencing pain’ can be helpfully identified by others in a shared world, neurally generated. Koch does not experience this fact for neither Koch nor pain exist in the mentalist sense (The Perfidious Brain).

Kant is partly right when he says that ‘Through this I... nothing is represented than a transcendental subject of the thoughts [or pain] = X’ (1933 p. 331). But Kant was expressly talking of a non-physical mind that thinks or has pain. Brains are physical; they do not think or have pain.

But how do we account for thinking about ourselves or our thoughts—meta-consciousness? How do we address knowing we are in pain? Daniel Wegner claims we (our selves) are ‘privy to our own experiences’. Certainly, we seem to think about things in the world *and* in ourselves, which are notionally quite different, because the object is in different places—outside and inside. But this is fallacious. There is no inner world: there is one physical world of which brains are part, and brain-signs are physical structures.

Our so-called inner life is interneural communication, whether we suppose we have pain (‘feeling’) or know we have it (‘reflection’). The construct of brain-sign is the biophysical marker plus the location of content in the only place it could be—the body *I am* (in Koch’s case a tooth). ‘Becoming aware of it’, so-called, *is* neurally communicating it. Saying ‘I have pain in my jaw’ follows the initial brain-sign and differs from it. The intent of saying it (before it is said) is not a mental reflection (or realisation) of being in pain. I.e. the supposition we think about ourselves as a self experiencing mental entity results from the theoretical structure of mentalism forced on us from infancy without question.¹² So, beyond the contortion of our face, another brain can identify and associate with (quasi-replicate as signification) our brain’s condition, when told, because brain-sign creates a common world. This is neither empathy as a dualist mental function, nor simulation theory as a mind state. But the latter is in-tune with brain-sign theory.¹³

Thus the philosophical notion of apperception (Descartes, Leibnitz, Kant)—the awareness of self being aware of the world—is superseded by the neuroscientific account of interneural communication. It occurs in our dog/cat (and other creatures) and is mistakenly supposed as mental selfhood. A creature ‘recognizing itself’ in a mirror (dolphins,

¹² As Will Storr puts it: ‘The controlled hallucination inside the silent, black vault of our skulls that we experience as reality is warped by faulty information. But because the distorted reality is the only reality we know, we just cannot see where it’s gone wrong’ (Storr 2019, p. 62).

¹³ See, e.g. Carruthers 2011, p. 224 and below.

¹¹ For example, *Embodiment, Enaction, and Culture* (Durt et al. 2017).

chimpanzees) is not recognising a mental self. It has an additional behavioural ability towards its body. Inside its head, it still signifies its interaction with others.

Here is another example. We are startled by an unbidden and very strange ‘thought’. And, as if peering into the depths, ‘we’ say to ‘ourselves’: ‘How could I have thought that bizarre thing?’ But just as the bizarre thought occurs spontaneously, so does the questioning ‘reflection’. However, it is not an introspective action by a mental I questioning a hidden (unconscious) interior producing bizarre thoughts.

The bizarre thought is a brain-sign construct from the brain’s immediate causal orientation. It is discontinuous with prior brain events. The brain constantly organizes and expresses its operation. Its processes are not controlled (or accessed) by a conscious or rational mind.¹⁴ But since the event is unrelated to prior causal orientations, the brain generates a communicative brain-sign noting its unrelated character. Because brain-signs function communicatively, the supposedly reflective event is, in principle, still communication with others. It is formed with the biophysical marker and a new causal orientation pointing towards the construction of the bizarre ‘thought’ by a fictive self/mind. The prior ‘thought’ is not coincident with it but, in these following brain-sign constructs, it hovers behind and before it temporarily. The notion of self-inquiry does not describe the brain’s operation, for the biophysical marker does no work beyond identifying this organism.

This description makes three points. (1) The biophysical marker is not a mental subject that can actively inquire into its mind. (2) Having identified the role of brain-sign, there is no question of ownership of thoughts by a mental subject. (3) As brain-sign, we are the brain’s explanation of its state which functions as the common world with others.

Certainly often no communication takes place. But the brain does not discriminate in its autonomic creation of brain-sign for it does not know what it is doing. Overproduction is characteristic of biological processes.

*

Now again it is important briefly to distinguish this account from others in the literature. It has become fairly common to state that the brain makes up stories, significantly influenced by the work of Michael Gazzaniga on split-brain patients (his ‘left brain interpreter’, e.g. 2011; also on consciousness, 2018). But in doing so the brain has no access to any kind of (transcendent) knowledge beyond the conditions of its own neural functioning. Here is an account from Nick Chater.

‘Our flow of conscious thought, including our explanations of our own and others’ behaviour, are creations of the moment, not reports of (or even speculations about)

a chain of inner mental events. Our mind is continually interpreting, justifying and making sense of our own behaviour, just as we make sense of the people around us, or characters in fiction’ (2018, p. 6). ‘There is no inner world’ (ibid., p. 8).

Brain-sign theory agrees the brain phenomenon is a creation of the moment and the brain generates an account of itself (further below). That there are no unconscious mental states is certain for none are conscious. But Chater’s mind and consciousness are given no biophysical credence, rendering ‘interpreting, justifying and making sense’ gratuitous. In particular, the conscious subject is completely absent. So to whom is ‘our’ behaviour made sense, and how does that happen? Surely sentences occur. ‘Looking at the page in front of me as I type, I have the feeling that I see words everywhere’ (ibid., p. 41). But this I has no scientific validation. It is simply assumed because it is how one talks (or types).

In a somewhat similar way, Dennett persists in not explaining consciousness whilst continuing to use the term. The issue is his fixation on language. He says, according to Patricia Churchland, that ‘without language, an animal is not conscious. That includes nonlinguistic humans.’ (2013, p. 204.) She demurs (after Panksepp 2010): ‘Being conscious enables the acquisition of language, not the other way round’ (ibid., p. 205).

Indeed, Dennett does not explain the conscious subject, the self. His view in 1991, repeated since, is that ‘If you think of yourself as a *center of narrative gravity*...your existence depends upon the persistence of that narrative...which could theoretically survive many switches of medium’ including (apparently) immortality if your name, the ‘center’, is in print (p. 430, emphasis added). You are simply information. Thus is consciousness wiped out by his inapt computer analogy.

In brain-sign theory, the biophysical marker replaces the conscious subject (there is no consciousness) and endorses Churchland’s claim about animals and nonlinguistic humans. For example, she proposes that ‘when my dog Duff sees me packing my suitcase and looks downcast...he is feeling the sadness of imminent separation...probably similar to my own’ (ibid., p. 249). The problem for Churchland is that ‘feeling sad’ is scientifically unexplained.

On the other hand, Peter Carruthers, denier of conscious agency, says that ‘While *selves* exist, they...should be thought to comprise all of the mental states...within the agent (the vast majority of which are unconscious)’ (2011, p. 380). (Note the conflict with Chater.) But by not eliminating consciousness and saying it is not epiphenomenal, he commits the self to vacuity.

Emotions, etc., Redesigned

Another element of brain-sign is categories-of-interaction. By contrast with world representations, these express the organism’s response to the world (cf. ‘[Foundation and Evolution](#)’).

¹⁴ For Kant, the ‘considered’ self of reflection (the ‘empirical’ I) is an ordinary representation, not a manifestation of the transcendental I of consciousness.

They also result from the brain's interpretation of its causal orientation and coincide with categories-of-the-world.

Mentalism, as a theory, states that if a lion appears before us we see it and feel fear. Brain-sign theory says we do not see it or feel fear. The brain and body react, but they do so purely as physical structures which includes signification to facilitate interneural communication about the world and the body's response to it. Whilst the complexity of language is a distinction between humans and other creatures, humans are not distinct from (many) other vertebrate creatures in the generation of brain-sign (Patricia Churchland and her dog).

As a high-level generality, categories-of-the-world are generated from activity in the neocortex; categories-of-interaction from the limbic system, the amygdala being a prime locus. The limbic system is older phylogenetically, but these two areas are influential on each other. That 'emotion interacts with cognition has become a fairly well accepted notion' states Luiz Pessoa (2013, p. 2).

Both categories derive from the causal orientation, so while an aim of neuroscience should be the identification of where and how brain-sign is located, it is equally necessary to determine where and how causal orientation occurs. The implication is that the brain's causal status has states of fast recurring reportability from which brain-sign is derived.¹⁵ This must be a condition with identifiable physical characteristics, by contrast with consciousness where neural assemblies are representationally intelligent (by unspecified means) and, as it were, glow, feel, sense and emote (cf. Tononi et al. 2016; Dehaene 2014—as above).

Koch says 'I feel the pain'. How can matter feel pain? Or be upset about it? Brain-sign theory resolves this because 'we are' a neural construct, not a mental event. States of (supposed) anger, adoration, envy, gratitude, serve interneural communication directly and are not merely behavioural attitudes. They do so because, as signs, they are a common brain occurrence, which is a tacit justification for their role. Antonio Damasio says: 'The thing to marvel at ... is the similarity not the difference [between organisms in emotional expression]' (1999, p. 53). However, his subsequent statement is that 'emotions produce quite reasonable behaviors' (ibid., p. 54—emphasis added). But the physical generator of an organism's behaviour is not an emotion; it is its causal orientation from which brain-sign derives. This renders brain science tractable. The associated behaviour fits into biologically determinable and classifiable structures.

Here is an example. Activated adrenalin (epinephrine) energises the body's condition in blood circulation for rapid response (Luo 2016, p. 353). So, is the sense of fear the sole result of adrenalin release? No. Adrenalin is also released in

vigorous exercise. Context is crucial. How is it determined? By the brain's causal orientation in brain-operational terms. The first is a threat; the second, exertion. Different categories-of-the-world arise with different categories-of-interaction—in the first case external activity the brain takes as threatening; for the jogger, the path ahead and leg ache. In other words, adrenalin's influence on causal orientation is qualified by circumstance. Claiming adrenalin causes fear is incorrect, as is seeing a threatening situation. Adrenalin is a constituent of causal orientation; seeing does not exist.¹⁶

Pessoa comments 'it may be time to stop describing concepts in terms of dichotomies [ref. Newell 1973; Kelso and Engström 2006] and to adopt a vocabulary that views concepts as complementary pairs that mutually define each other and, critically, do not exclude each other' (2013, p. 5). Therefore 'I believe it is more fruitful to describe mental phenomena in terms of cognitive-emotional interactions' (ibid.).

Brain-sign theory discards mentalist terminology (i.e. Pessoa's 'descriptive' usage of cognition-emotion) locating brain-signs in neural structures and conditions derived from the causal orientation. The point is not that this can be done today; rather it offers a constructive route in grounding future brain science.

An example is, 'disappointment' resulting from an event or situation not matching hopes or wishes. The person, the supposed self, feels 'let down' or failed in 'what was wanted'. The brain-sign reconstruction emphasises causal orientation (immediate or subsequent) as brain networks directed at a biophysically describable relation toward the world (generating control, co-opting allies, endorsement of actions). Organisms are continually balked in such events, to a greater or lesser degree. They adapt their networks accordingly, e.g. altering their instructions or aborting the activation. What does not happen is human subjects pondering the situation and making mental decisions (cf. Carruthers). So, while pain/pleasure or love/hate accounts have been neurally generated as simpler explanations than brain analysis, they are inadequate for science (e.g. Freudian psychology, and contra Solms 2018). Categories-of-interaction related to causal orientation offers a plausible investigative route.

As mentioned, the motivations people are deemed to have are non-scientific categories, as are Aristotelian 'purposes'. The physical universe has no purposes or motivations (contra Pessoa, chapter 6; Berridge 2018). The current concern for mental health as distinct from physical health (daily expressed in the news) is as scientifically unhelpful as phlogiston before molecular chemistry.

*

The terms 'exteroception' and 'interoception' have become current: the first, perception of the external world; the second

¹⁵ An indicative example is *chronostasis* resulting from the rapid movement of the eyes (the saccade) to a new target resulting in the *stopped-clock illusion*. The second hand appears to have a longer immediate delay before moving than subsequently because the brain anticipates the saccade. See further Knoll et al. (2013).

¹⁶ For recent discussion of constitution and causality, see, e.g. Kästner and Anderson (2018).

perception of inner states of the body. Obviously, the terms presume the capacity for ‘perception of...’ As Cynthia Price and Carole Hooven put it: ‘What becomes conscious, i.e. interoceptive awareness, involves the processing of inner sensations so they become available to conscious awareness’ (2018).

Erik Ceunen et al. (2016) claim that ‘‘Interoception’ is a concept which relates to a...wide range of health...and psychological aspects of human life, playing a role in every individual.... A cursory glance at the literature is sufficient to see ... a vast range of subjects.’ The authors generate a list of ‘inner’ conditions including: medically defined symptoms, emotions in general, decision making, and subjective time perception. Pain is particularly significant. They quote The International Association for the Study of Pain (IASP). Pain is ‘An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage’ (Merskey and Bogduk 1994). This makes no attempt at science. Giving this diversity a technical name does not solve the mind-body problem.

The authors say that ‘The only thing determining whether something is interoceptive is whether it contributes to the subjective perception of body state.’ So whether there is actual damage, in the case of pain, is irrelevant to interoception if it is felt. For example, ‘The findings [of McGrath et al. 2013] suggest that at least for major depression, and perhaps for mood and other disorders, there are two subgroups of patients...one...with an overactive anterior insula, the other... with an underactive anterior insula. These two...neurological biomarker patterns...are suggestive respectively of accurate and inaccurate perceivers.... Regardless of the accuracy with which individuals with mood and anxiety disorders can perceive sensory homeostatic afferent feedback, individuals with such disorders excessively rely on sources other than the actual bottom-up homeostatic pathways, giving more weight to maladaptive cognitive-emotional schemes of interpretation (Paulus and Stein 2010).’ I.e. over- and underactive neural conditions generate ‘biased’ accounts.

This is supported by the authors’ comment that ‘Seth (2013) proposes that interoception...is not just passive, bottom-up processing, but...also involves active top-down activation to make predictions of the causes of sensory input. His view is...that perception is a process of not only afferent feedback, but also of predictions, and ultimately the integration of both, resulting in prediction errors.’ An expanded constructivist account (cf. Koenderink and Krueger 2017), now concerning ‘inner’ states.

Brain-sign theory contributes twofold. There are no subjective phenomenological states, and there is a clear discrimination of the world as represented (including what is termed imagined and remembered), and a response to it. So, ‘pain’ which is *not* caused by body damage yet ‘felt’ is categories-of-the-world *and* categories-of-interaction. The pain ‘felt’ is

bodily, though an operational misnomer—i.e. how the brain categorises it as brain-sign. Major depression and mood disorders (categories-of-interaction) are neurologically communicated representations. Which is why they ‘play a role for every individual,’ as the authors say, and not because mental states exist or influence behaviour.

Language Redesigned

Language has been a great mystery. How does it raise us beyond other creatures? How can words change minds or influence actions? But at the heart of these questions is a profound mistake.

Brain-sign language is not language as generally supposed, i.e. that we (mentally) understand words and sentences and act upon them. Understanding is a prescientific concept, which dissolves a central plank of Kant’s doctrine. But before addressing the mistake, the principles of the brain-sign theory approach are outlined.

Because the universe is physical, a scientific theory needs to determine the physical characteristics of the linguistic process. Language is not a phenomenon of mind. It is an evolved means for one organism to change the causal orientation of another (*contra* Chomsky, e.g. 2016; Dehaene 2014, p. 250). This is effected by the generation and reception of compression waves between brains from/to vocal mechanism and ears. In the case of humans, it also entails transmission by electromagnetic radiation via the eyes (‘reading’). Physical transmission is a basic inter-organism practice. Many creatures sign by molecular transmission for location marking. Urine spray is customary for dogs. So language has no transcendent characteristics. Complexity is the feature language offers, resulting from associated brain capacity.

The foundational principle is that language is ‘learned’ and maintained, not as words and sentences, but as neural elements of causal orientation. That is, words are linked (as brain-signs) with behavioural actions in relation to the organism’s engagement with the world. For example, the word ‘apple’ activates behavioural networks associated with apples. When one brain (A) executes the transmission of compression waves to another (B), the word ‘apple’ occurs in the latter’s brain as brain-sign. But the causal path for the word to become the brain-sign occurrence results from the association in B’s brain with its causal orientation towards apples. In other words, brains do not have dictionaries of words with meaning. B’s brain-sign of ‘apple’ is a signification of causal networks in B’s brain. The word ‘apple’, which A and B then hold in common, is a joint association of causal possibilities (actions in the world) associated with apples. Words (so-called) are neural signifiers of causal orientations, i.e. subsets of the totality of possibilities, and not added mental conditions that are understood and can affect actions. How the brain generates brain-sign in this situation is not yet established. However, its operational function is on the same grounds as other brain-sign elements.

Thus, in training an infant to associate a word with an object, the adult attempts to generate in the child the same neural operation he/she has. However, the adult has no access to the nature of his/her own operative neural structure because that is not revealed by the brain in its operation. Brain-sign itself, as said, has no causal impact on the organism's own actions. It is purely a sign.

What *seems* to happen is that the adult points to an object and says its name. By that method, in an opaque way, the adult forces the word into the infant's head for each occasion the object is before it. But while this is an automatic method of the concerned parent (as 'supposed' by him/her), what happens is more complex. When the adult engages with an object with its brain causally orientated towards it, the neurally interpreted name as brain-sign can also occur. These two brain-sign elements (object and word) coexist. This elaborate neural event is to be installed in the child. The child's brain is being trained to associate its causal orientation towards the object, from which its brain-sign of the object is derived, with its causal orientation towards the compression waves impact by which the name will become a linked neural structure. After the structure is built, and when activated by compression waves from an external source, the name will occur in the child as a brain-sign evoking the causal neural structures associated with the object. This replaces the notion of meaning. The 'sense' we know what a word means *is* the communicative act (cf. 'perception', 'Perception Redesigned'). The occurrent word does not reference the world object.

In summary, the causal orientation of one brain can alter the causal orientation of another, which is signified by the coincidence of brain-signs (words and objects) from those causal orientations. The word 'is', for example, evokes that to which causal orientation is possible, not the mysterious notion *Being* which inspired Heidegger ((1927), 1962).¹⁷ Searching for mental states of language is a misapplied effort. Of course, neuroscience should be concerned with the where and how brain-sign is manufactured; but as a sign, it is not a gift of knowledge. Talking to each does not concern world reality; the words signify brain causality which is an 'invisible' reality.

As the infant grows, the causal properties of its brain develop, so associations with what a word (and its influence within sentences) signifies vastly expand in its potential action capability. This is inaccurately termed greater knowledge or wisdom. Reading a story (electromagnetic communication), while apparently an enjoyable occupation for the child (categories-of-interaction), is actually an engagement of the brain's causal orientations. No human gets to the raw processes of the brain because the brain builds its causal functionality and then signifies it for communication.

*

The Preface to W.V.O. Quine's book *Word and Object* exactly reflects brain-sign functioning. 'Language is a social

¹⁷ Heidegger says: 'It is said that 'Being' is the most universal and emptiest of concepts.' *Being and Time* (p. 2).

art.... There is no justification for collating linguistic meanings unless in terms of men's dispositions to respond overtly to socially observable stimulations. An effect of recognizing this limitation is that the enterprise of translation is found to be involved in a systematic indeterminacy' (1960, p. ix). The point is not that 'dispositions to respond' or 'stimulations' are not the base conditions for 'indeterminacy'. It is that brain communication by signification depends upon neural structures (causal orientations) which are different in individuals, particularly across languages. They serve an often communicative adequacy.

Summary

Brain-sign theory offers an account accessible as science which mentalism does not.

Subjectivity is replaced by biology enabling collective action amongst organisms. That behaviour exists in comparatively primitive organisms but reaches its complexity-zenith and flexibility in humans, who co-signify the world in which joint action can occur. It is not an input to the brain, but a profile of action in the world held in common. Many writers emphasise the ability of humans to communicate, but their approach is dogged by the history of mind as a private individualised condition.

The theory has two major impacts. (1) The structures and states of neural causal orientations, from which brain-signs derive, require a descriptive vocabulary identifying resulting behaviour. This is a major task since there is no likelihood of one-to-one correspondence, and causal brain states are highly complex. However the assumptions of mentalism obscure this crucial activity. Indeed, it is likely that the current identification of active brain areas associated with mental function is often associated with causal orientations. However, brain-signs are brain explanations of its operation which may, therefore, aid neuroscientific investigation. (2) The theory alters human self-conception, including denial of knowledge. Humans neither know nor believe anything. Neural structures and states determine the organism's action in the world, not word formations or feelings. This way of viewing human life is a scientific upgrade from psychology. While radical as a proposal—though individuals have proffered adjacent hypotheses—the function of brain-sign completes a scientific account that potentially could improve human decision making and behaviour, so influencing adaptation and survival. These will be long term developments.

Brain-sign is a scientific hypothesis. But the actuality of its implementation cannot be known (or explained). As Galileo said, 'There is not a single effect in nature... such that the most ingenious theorist can arrive at a complete understanding of it.'¹⁸ Brain-sign theory is why.

¹⁸ Quoted from Chomsky 2016, p. 91.

Compliance with Ethical Standards

Conflict of Interest The author declares that he has no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anderson, M. L. (2014). *After phrenology: neural reuse and the interactive brain*. A Bradford book. MIT Press.
- Baars, B. J. (1997). *In the theater of consciousness*. New York: Oxford University Press.
- Bargh, J. A. (2005). Demystifying the will: toward demystifying the nonconscious control of social behavior. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 37–58). Oxford University Press.
- Barrington, R., Gunton, M., Barton, M., Grey, I, Hugh-Jones, T. (2014) Why a tiny male creates a big-wheel masterpiece. In *Life story*, BBC Books.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74(5), 1252–1265.
- Berridge, K. C. (2018). Evolving concepts of emotion and motivation. *Frontiers in Psychology*, 9, 1647. <https://doi.org/10.3389/fpsyg.2018.01647>.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227–287. <https://doi.org/10.1017/S0140525X00038188>.
- Block, N. (2015). Consciousness, big science, and conceptual clarity. In G. Marcus & J. Freeman (Eds.), *The future of the brain* (pp. 161–176). Princeton University Press.
- Butler, R. (2012). The arts of deception: mimicry and camouflage. Monga Bay Rainforest Diversity. <https://rainforests.mongabay.com/0306.htm>.
- Buzsáki, G. (2019). *The brain from inside out*. Oxford University Press.
- Carruthers, P. (2011). *The opacity of mind: an integrative theory of self-knowledge*. Oxford University Press.
- Carruthers, P. (2018). There is no such thing as conscious thought. *Scientific American*. <https://www.scientificamerican.com/article/there-is-no-such-thing-as-conscious-thought/>.
- Ceunen, E., Vlaeyen, J. W. S., & Diest, I. V. (2016). On the origins of interoception. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2016.00743>.
- Chabris, C. F., & Simons, D. J. (2010). *The invisible Gorilla: how our intuitions deceive us*. Crown Publishers.
- Chalmers, D. J. (1996). *The conscious mind: in search of a fundamental theory*. Oxford University Press.
- Chater, N. (2018). *The mind is flat: the illusion of mental depth and the improvised mind*. Allen Lane.
- Chomsky, N. (2016). *What kind of creatures are we?* Columbia University Press.
- Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, LXXVIII. <https://doi.org/10.5840/jphil198178268>.
- Churchland, P. S. (2013). *Touching a nerve: our brain, our selves*. W.W. Norton.
- Clapson, P. (2016). Communication without consciousness: the theory of brain-sign. *Activitas Nervosa Superior.*, 58, 84–107. <https://doi.org/10.1007/BF03379739A>.
- Cobb, M. (2020). *The idea of the brain: a history*. Profile Books Ltd..
- Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, 6(2), 119–126.
- Damasio, A. (1999). *The feeling of what happens*. William Heineman.
- Dawkins, R., & Krebs, J. R. (1978). *Animal signals: information or manipulation? Behavioral ecology: an evolutionary approach* (Vol. 2, pp. 282–309). Sinauer Associates, Inc..
- Dehaene, S. (2014). *Consciousness and the brain: deciphering how the brain codes thoughts*. Viking Penguin.
- Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70, 200–227. <https://doi.org/10.1016/j.neuron.2011.03.018>.
- Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science*, 358(6362), 486–492. <https://doi.org/10.1126/science.aan8871>.
- Dennett, D. (1991). *Consciousness explained*. Little, Brown and Company.
- Dennett, D. (2017). *From Bacteria to Bach and Back*. Allen Lane.
- Donald, M. (2001). *A mind so rare: the evolution of human consciousness*. New York: Norton.
- Durt, C., Fuchs, T., & Tewes, C. (2017). *Embodiment, enaction, and culture: investigating the constitution of the shared world*. MIT Press.
- Frässle, S. J., Sommer, J., Jansen, A., Naber, M., & Einhauser, W. (2014). Binocular rivalry: frontal activity relates to inspection and action but not to perception. *Journal of Neuroscience*, 34(5), 1738–1747. <https://doi.org/10.1523/JNEUROSCI.4403-13.2014>.
- Frith, C. D., & Rees, G. (2017). Methodologies for identifying the neural correlates of consciousness. In M. Schneider & M. Velmans (Eds.), *The Blackwell companion to consciousness* (2nd ed., pp. 591–606). Blackwell Publishers.
- Heidegger, M. ((1927), 1962). Being and time. Trans. J. Macquarrie and E. Robinson. Basil Blackwell.
- Hume, D. (1739). *A treatise of human nature*. London: John Noon.
- Husserl, E. ((1954) 1989). *The crisis of European sciences and transcendental philosophy: an introduction to phenomenological philosophy*. Trans. D. Carr. North Western University Press.
- James, W. (1890). *Principles of psychology*. New York: Holt.
- Kant, I. ((1787), 1933). *Immanuel Kant's critique of pure reason*. Trans. N.K. Smith. The MacMillan Press.
- Kästner, L., & Anderson, L. M. (2018). Intervening into mechanisms: prospects and challenges. *Philosophy Compass*. <https://doi.org/10.1111/phc3.12546>.
- Kelso, J. A., & Engström, D. A. (2006). *The complementary nature*. Cambridge: MIT Press.
- Knoll, J., Morrone, M., & Bremmer, F. (2013). Spatio-temporal topography of saccadic overestimation of time. *Vision Research*, 83, 56–65. <https://doi.org/10.1016/j.visres.2013.02.013>.
- Koch, C. (2018). What is consciousness? *Nature*, 557, S8–S12.
- Koenderink, J., & Krueger, J. (2017). Rationality, perception and the all-seeing eye. *Psychonomic Bulletin and Review*. <https://doi.org/10.1162/biot.2009.4.2.167>.
- Kuhn, T. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.

- Lamme, V. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 15(8), 365–373. [https://doi.org/10.1016/s1364-6613\(02\)00013-x](https://doi.org/10.1016/s1364-6613(02)00013-x).
- Laureys, S., Gosseries, O., & Tononi, G. (2016). *The neurology of consciousness: cognitive neuroscience and neuropathology*. Elsevier.
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of neural activity (readiness potential). The unconscious initiation of a freely voluntary act. *Brain*, 106, 623–642. <https://doi.org/10.1007/978-1-4612->
- Luo, L. (2016). *Principles of neurobiology*. Garland Science, Taylor and Francis.
- McFarland, D. (1989). *Problems of animal behaviour*. UK: Longman Scientific and Technical.
- McGrath, C. L., Kelley, M. E., Holtzheimer, P. E., Dunlop, B. W., Craighead, W. E., Franco, A. R., et al. (2013). Toward a neuroimaging treatment selection biomarker for major depressive disorder. *JAMA Psychiatry*, 70, 821–829. <https://doi.org/10.1001/jamapsychiatry.2013.143>.
- Melnikoff, D.E., Bargh, J.A. (2018). The mythical number two. Trends in cognitive science: Cell Press Reviews. doi: <https://doi.org/10.1016/j.tics.2018.02.001>.
- Merskey, H. E., & Bogduk, N. (1994). *Classification of chronic pain, IASP task force on taxonomy*. Seattle: International Association for the Study of Pain Press.
- Moritz, R. F. A., Kirchner, W. H., & Crewe, R. M. (1991). Chemical camouflage of the death's head hawkmoth (*Acherontia atropos* L.) in honeybee colonies. *Naturwissenschaften*, 78, 179–182. <https://doi.org/10.1007/BF01136209>.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435–450. <https://doi.org/10.2307/2183914>.
- Newell, A. (1973). *You can't play 20 questions with nature and win: projective comments on the papers of this symposium and win*. Pittsburgh: Carnegie Mellon University, Department of Computer Science.
- Nietzsche, F. ((1887) 1974). *The gay science*. Trans W. Kaufmann. Vintage Books.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, 84, 231–259. <https://doi.org/10.1037/0033-295X.84.3.231>.
- Noë, A. (2004). *Action in perception*. MIT Press.
- Nørretranders, T. ((1991), 1998). *The user illusion: cutting consciousness down to size*. Trans. J. Sydenham. Viking Penguin.
- Oakley, D. A., & Halligan, P. W. (2017). Chasing the rainbow: the non-conscious nature of being. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2017.01924>.
- Panksepp, J. (2010). Affective consciousness in animals: perspectives on dimensional and primary processes emotion approaches. *Proceedings of the Royal Society*, 277(1696), 2905–2907.
- Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Structure & Function*, 214, 451–463. <https://doi.org/10.1007/s00429-010-0258-9>.
- Pennartz, C. M. A., Farisco, M., & Evers, K. (2019). Indicators and criteria of consciousness in animals and intelligent machines: an inside-out approach. *Frontiers in Systems Neuroscience*. <https://doi.org/10.3389/fnsys.2019.00025>.
- Pessoa, L. (2013). *The cognitive-emotional brain: from interaction to integration*. MIT Press.
- Poeppel, D. (2015). The neurobiology of language. In The future of the brain. G. Marcus, J. Freeman (Eds). Princeton University Press.
- Price, C., & Hooven, C. (2018). Interoceptive awareness skills for emotion regulation: Theory and approach of mindful awareness in body-oriented therapy (MABT). *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2018.00798>.
- Quine, W. V. O. (1960). *Word and object*. MIT Press.
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17, 565–573. <https://doi.org/10.1016/j.tics.2013.09.007>.
- Schatzki, T. R. (1992). Being, the clearing, and realism. In H. L. Dreyfus & H. Hall (Eds.), *Heidegger: a critical reader* (pp. 81–98). Blackwell Publishers.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentive blindness for dynamic events. *Perception*, 28(9), 1059–1074. <https://doi.org/10.1068/p281059>.
- Solms, M. (2018). The neurobiological underpinnings of psychoanalytic theory and therapy. *Frontiers in Behavioural Neuroscience*. <https://doi.org/10.3389/fnbeh.2018.00294>.
- Stewart, J. (2014). Foundational Issues in Enaction as a Paradigm for Cognitive Science. In J. Stewart, O. Gapenne, E. A. Di Paolo (Eds.), *Enaction: Toward a Paradigm for Cognitive Science*. MIT Press.
- Storr, W. (2019). *The science of storytelling*. William Collins.
- Strawson, P. (2017). One hundred years of consciousness ('a long training in absurdity'). <https://podcasts.ox.ac.uk/one-hundred-years-consciousness-long-training-absurdity>.
- Tomasello, M. (2014). The ultra-social animal. *European Journal of Social Psychology*, 44, 187–194. <https://doi.org/10.1002/ejsp.2015/ejsp>.
- Tononi, G., Melanie, B., Gosseries, O., & Laureys, S. (2016). In S. Laureys, O. Gosseries, & G. Tononi (Eds.), *The neurology of consciousness: an overview* (2nd ed.). Elsevier.
- Varela, F. J., Thompson, E., Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.
- Wegner, D. (2002). *The illusion of conscious will*. MIT Press.
- Whyte, L. L. (1960). *The unconscious before Freud*. New York: Basic Books.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.