‘Trying to find consciousness in the brain’

That there is an intimate relationship between our consciousness and the functioning of our brains was not a novel idea even when Hippocrates famously asserted that ‘from the brain, and from the brain only, arise our pleasures, joys, laughter and jests, as well as our sorrows, pains, grief and tears’ (quoted in Spillane, 1981). The evidence—from everyday observations such as the effect of a bump on the head to the most sophisticated neuroscientific experiments—seems overwhelmingly to support the belief that the quantity, distribution and spatio-temporal patterns of neural activity in the brain are somehow connected with the level and content of consciousness. So much is clear. More opaque is the precise nature of the connexion.

There are different ways of interpreting the broad correlation between neural activity and consciousness. Nowadays, by far the most common view is that experiences are *identical* to the relevant activity. Unfortunately, this theory—which philosophers have examined in minute detail in many thousands of articles since U. T. Place first advanced a modern formulation in 1956 (Place, 1956; Borst, 1970)—is not at all straightforward. It has at least two major variants. According to the first, neural activity simply *is* conscious experience—end of story. Unfortunately, nerve impulses do not seem like experiences, such as awareness of yellow or of pain; and the least one might expect is that something should seem like itself. This has prompted the suggestion that neural activity and conscious experiences are two different *aspects* of what are, essentially, the same event. This interpretation, often called ‘property dualism’, to contrast it with the ‘substance dualism’ of traditional Cartesian thought, is very popular with both philosophers and neuroscientists.

For example, John Searle, one of the most prominent philosophers of consciousness, has advanced the dual aspect theory but tried to make it more palatable as a dual *level* theory (Searle, 1983). Neural activity and conscious experience, he argues, are analogous, respectively, to molecules of H₂O and liquid water. While water has properties such as shininess and dampness that molecules of H₂O do not possess, the latter are clearly identical to the former. So it is with neural activity and subjective experiences. The analogy is, of course, flawed: we cannot compare the relationship between experiences and their putative physical basis with the relationship between different ways of experiencing the same thing, or between the appearances of a material such as water observed at different levels (Tallis, 2004). Indeed, we cannot help ourselves to the notions of ‘levels’, ‘appearances’ or ‘observations’ without presupposing consciousness to underpin them.

Others have ignored the need to explain how nerve impulses and subjective experiences have somehow to be ‘the same and..."
not the same’ and have stuck doggedly to single aspect identity theories. In order to make this more convincing, they have sidelined, or even (as in the case of Daniel Dennett, their most prominent exponent) denied, the reality of ‘subjective experiences’ (Dennett, 1991). Worrying about ‘qualia’—actually experienced contents of consciousness—is a hangover from a pre-scientific psychology. A mature account of mind will by-pass these pseudo-entities and, instead, focus on their functional equivalents, i.e. the patterns of causal relationships between neural inputs and outputs. There are inputs of sensory stimuli, outputs that are dispositions to behaviour or ‘reactive dispositions’, and in between there are ‘discriminative states’, i.e. neural activities whose rule-governed behaviour ensures that the right (i.e. adaptive) outputs result from the appropriate inputs.

Many people think this is cheating. David Chalmers (Chalmers, 1996) argues that any theory of consciousness worth the name should take account not only of what he calls the ‘psychological concept of mind’ as a causal or explanatory basis for behaviour (which might be taken care of by functionalist accounts) but also ‘the phenomenal concept’ of mind, in which mental states are consciously experienced. This has led him to a much-discussed theory which combines property dualism with functionalism (Chalmers, 1996). Consciousness, he claims (p. 247), arises in virtue of ‘the functional organization of the brain’ which is ‘best understood as the abstract pattern of causal interaction between various parts of a system’. The appeal to patterns, a fortiori abstract ones, seems as illegitimate as Searle’s appeal to levels, appearances and observations. What is more, his attempt to explain how phenomenal consciousness arises, by seeing it as the necessary product of ‘fine-grained functional organization’, leads him to accepting that anything that has such a functional organization will have consciousness. This is because the functional organization is a physical realization of information and, where there is a physical realization, there is a phenomenal realization. In the last analysis, ‘information’ is simply ‘a difference that makes a difference’ (p. 281). This extended idea of ‘information’ forces him to concede that where there are such differences, there is physical information; and where there is information, as in thermostats, there is experience. ‘One can find information states in a rock, when it expands and contracts, for example, or even in the different states of an electron’. So there will be experience associated with a rock or an electron. (Chalmers, 1966, p. 297).

The path to much lunacy in neurophilosophy is paved with the misuse of terms, and ‘information’ is particularly likely to be implicated. ‘Information’ has a narrow, highly specialized sense in communication engineering, where it has nothing to do with semantic information or with meaning. Neurophilosophers use ‘information’ in the engineering sense while exploiting the explanatory force that comes from its everyday sense of something that passes between a person who consciously informs someone of something, and another who is consequently consciously informed.

The theories of Searle, Dennett and Chalmers (in ascending order of unsatisfactoriness) show how intractable are the challenges faced by identity theorists. The fundamental problem is that there seems to be nothing in neural activity, understood as events in the material world and subject to the latter’s laws, to explain the existence of a ‘what it is like’ to be a creature with conscious experiences (Nagel, 1974). What is more, on close examination, subjective consciousness proves to have some very awkward properties: ‘aboutness’ or intensionality; value (toothache and orgasms, success and humiliation, are differently valued); an ineliminable indexicality (pitching an egocentric space with ‘here’, ‘now’, ‘I’ etc., in a material space that lacks such ‘personal’ coordinates); and unity amid diversity (see Tallis, 2005, for a recent summary of these problems). Intentionality is especially resistant to physical explanation. While the relationship between light energy originating from an object impinging on the retina and nerve impulses in the visual system can be understood in terms of causal mechanisms seen throughout nature, the fact that those nerve impulses, or some of them, are ‘about’ the object—that, as it were, they ‘reach out’ or reach ‘causally upstream’ to it—cannot be so understood. We comprehend how the light comes into the brain but not how the gaze looks out of it. Notwithstanding the assertions, assumptions and ingenious formulations of many neurophilosophers, the inward causal arrow cannot be turned round and made into an outward intentional, referential or ‘semantic’ link.

The neurophilosophy of consciousness, in short, is in a mess. Some writers such as Colin McGinn (1989) conclude that we shall never understand the transformation of ‘the water of the physical brain into the wine of consciousness’. And this seems plausible not for the reason McGinn gives that we are not biologically equipped to do it, but for a deeper reason. Science, which grows out of subjective consciousness, cannot in turn enclose it: consciousness is therefore bound to be a mystery—in Gabriel Marcel’s sense of the term—because it is ‘a problem which encroaches upon its own data and so invades them and so is transcended qua problem’ (Marcel, 1965). That which we are trying to explain includes our explanations, and ourselves, the would-be explainers.

Many of the problems of neurophilosophy are acknowledged by Adam Zeman. Anyone new to the field should begin with his beautifully written Consciousness. A User’s Guide, an extraordinarily helpful account of the relevant neuroscience and of the key philosophical arguments. I know of no better general exposition of the current state of our knowledge of how the brain works—a perfect complement to his brilliant but more narrowly focused invited recent review in Brain (Zeman, 2001). What is more, he has no wares to sell: his viewpoint is one of fully informed uncertainty. He provides an incomparably clear guide to theories put forward by others, including those of Gerald Edelman and of Christof Koch whose most recent books are also under review.

Edelman, in contrast, has an only intermittent awareness of the philosophical problems of the identity theory. Nonetheless, he hopes in Wider than the Sky. The Phenomenal Gift of Consciousness, ‘to disenthral those who believe that the subject of consciousness is...necessarily mysterious’ (p. 145). His
conviction ‘that consciousness is susceptible to scientific study has been supported’, he says, ‘by a sharp increase in the number of publications on the subject’ (p. xi). Koch is equally upbeat in *The Quest for Consciousness. A Neurobiological Approach*. ‘We live at a unique point in the history of science. The technology to discover and characterize how the subjective mind emerges out of the objective brain is within our reach. The next years will prove decisive’ (p. 314). Such optimism would not be possible without the (largely unconscious) habit among neuroscientists of anthropomorphism. Like the philosophers, Edelman and Koch credit brains, bits of brains, neural tracts, large and small circuits and even individual nerve impulses with doing things that strictly only conscious human beings, or machines working in conjunction with conscious humans, can do. Koch attributes ‘explicit coding’ to individual neurons and claims that ‘explicit representation of [a] stimulus attribute is a set of neurons that ‘detects’ that feature without much further processing’ (p. 306). Edelman, who is slightly more circumspect, and critical of the lax use of terms such as ‘representation’ and ‘calculation’, nonetheless describes dynamic structures as carrying out ‘global mapping’ whose function is ‘first to sample the world of signals by movement and attention and then to categorize these signals’ (p. 49). Both cross the explanatory gap, as many philosophers do, by combining machine-like accounts of the operation of the conscious mind with machino-morphic and anthropomorphic accounts of what goes on in brains. (Tallis, 2004).

One of the most telling problems of neurophilosophy is that, as Zeman points out, ‘while consciousness is bound up with the brain, not all the activity occurring in the brain is conscious’. This focuses the neurophilosophical question more narrowly: what is it that distinguishes conscious from unconscious brain activity? The time-honoured answers have been: location; quantity; and pattern. None of these seems very satisfactory. Take location. Why should physico-chemical activity in one part of the brain be conscious, or constitute consciousness, while very similar activity in another part is physico-chemical activity period? As for quantity, what intuitive appeal this has is derived from an illegitimate transfer of the notion of a perceptual threshold, which presupposes the existence of consciousness, or its possibility, the basis of consciousness itself. The fact that a smaller stimulus is not felt, while a bigger one is, guides the intuition that higher intensity of activity centrally will be conscious, consciousness or associated with consciousness (the ambiguity is revealing) while a lower intensity will not. Of course this metaphor of an exceeded threshold—that ‘more’ means ‘more likely to be conscious’—presupposes a background of an established consciousness which has a threshold that may be exceeded. And this is precisely the thing that has to be explained. The appeal to patterns is doubly unsatisfactory. There is no reason why one pattern should be associated with consciousness and another not, given that we are conscious of a multitude of patterns. Secondly, patterns are not intrinsic to material events: they have to be extracted from material events, by a consciousness (Tallis, 1999). At any rate, it is not clear where, or to what, a pattern arising out of a multitude of nerve impulses dispersed in space and time, and amongst other impulses that are not part of the pattern, is available to constitute a unity.

The unsatisfactoriness of location, intensity and pattern as explaining the difference between activity in the brain that is, and activity that is not, consciousness may be one of the drivers behind theories such as those of Edelman and Koch that shift between location, intensity and pattern. Another driver is the need to reconcile the unity of the moment of consciousness with the fact that, at any given conscious moment, we are aware of a multitude of distinct components. My instantaneous sensory field is unified, while I distinguish the colour, shape, velocity and location of seen objects and separately experience sounds, sights and smells. Within the single moment of consciousness, I distinguish bodily experiences and experiences of things outside of my body; memories and current experiences; fantasies and reality; objects of thought and objects of experience. How are they are they integrated while at the same time differentiated? How do we get merging without mushing?

Edelman’s Theory of Neural Group Selection purports to deal both with qualia and with the unitary nature of consciousness. Consciousness, he says, is not the property of neural activity in particular areas of the brain but of the ‘dynamic core’—‘a system of interactions, figured mainly in the thalamocortical system, which behaves as a functional cluster’ (p. 156). A functional cluster is ‘a system or part of a system that interacts mainly with itself’ (p. 159). A key notion is that of ‘reentry’: an evolving process of ‘recursive signalling across massively parallel reciprocal fibres’ connecting brain maps. Re-entry supposedly explains both qualia and the globality of consciousness. Re-entry makes neural activity aware of itself, converting ‘the signals from the world and the brain into a “phenomenal transform”—into what it is like to be that conscious animal’ (pp. 77–78). More specifically, ‘conscious processes arise from the reentrant interactions between value-category memory systems that are largely present in the more anterior parts of the thalamocortical system and the more posterior systems that carry out perceptual categorization’. The idea that mutual feedback between its component (unconscious) neural activity systems makes the dynamic core conscious is rooted in a rather naive intuition: that re-entry is a kind of neural introspection. Manifestly it is not: re-entrant activity is no different in kind from other activity in other systems. It is only from the viewpoint of a conscious external observer that re-entry amounts to self-visiting. Edelman’s notion that ‘higher-level’ consciousness—‘which allows its possessor to be conscious of being conscious, and to have a socially defined nameable self, and to have a concept of past and future’ (p. 116) and arises ‘by evolution of an additional reentrant capability’—shows how seriously and literally he takes his naive metaphor.

Not that he is very comfortable with his identification of qualia with neural activity. As he admits, it raises the spectre of epiphenomenalism, such that conscious experiences only play on the surface of causal processes, and ‘Consciousness is a
property of neural processes and cannot itself act causally in the world” (p. 141). If this were true, and it did not have any power to deflect the course of events and bring anything about, it would be difficult to understand why it evolved in the first place. This is an important failing for a neuroscientist who sees his work as ‘completing’ Darwin’s ‘program’ by developing ‘a view of consciousness as a product of evolution’ (p. 2).

Ever more refined mechanisms, one would have thought, would be the only survival tools worth having in a causally closed universe. Nor is re-entry very impressive as ‘a key process assuring …conscious integration’ (p. 179) or as ‘a global brain theory explaining diversity and integration in the central nervous system’. It may seem so only because he sees re-entry as a flow of ‘information’ (that word again!) between different areas. As Zeman explains it, these ‘allow local cortical circuits to perform their specialised ‘segmented’ functions while at the same time entering into a ‘unifying dialogue’’ (p. 289). I cannot see how the functional cluster can be both unified and differentiated. It seems to be able to be both, only because an outside observer can adopt a viewpoint that can focus electively on its unity or its multiplicity. In the absence of such a gaze, it is, of course, neither the one nor the other. Needless to say, Edelman’s account does not even hint at an explanation of how different facets of consciousness—memory, perception, emotion—are differentiated. Any explanation of how they are brought together makes it more difficult to imagine how they are also kept tidily apart, though the introspective reality of the unity and multiplicity of the moment of consciousness cannot be denied.

This unity-across-diversity of consciousness is a central preoccupation of Koch who, with Francis Crick, has developed an alternative account (which I shall call the KC hypothesis) of what differentiates neural activity that is associated with consciousness from neural activity that is not, and of what holds the innumerable components of consciousness together. Koch accepts that the large number of potential states accessible to the conscious mind—at any given moment, I can see a sunset, remember past sunsets and think about what certain poets said about them—necessitates the tight interaction of very large neural assemblies reaching right across the brain. However, he argues, such non-localizing explanations fail to explain why some kinds of widespread activity within the brain generate behaviours associated with consciousness and others do not. Koch is less interested in what he calls ‘the enabling factors for consciousness’—tonic conditions and systems that are needed for any form of consciousness to occur at all—than the specific factors required for any particular conscious percept. Enabling factors go beyond the brain, and include having an adequate blood supply. Within the brain, they encompass the activity of the brainstem reticular formation and the non-specific thalamic nuclei. Whether it is legitimate to think of ‘consciousness as such’ independently of specific contents—so that one can be conscious but of nothing in particular—is doubtful. (Level and content seem, to me, to be not so easily separable.) This does, however, clear the way for an approach to identifying the ‘circumscribed’ neural correlates of consciousness (NCCs) that Koch and Crick have been seeking over the last decade or so. The KC hypothesis is that the NCC at any moment correspond to the activity of shifting coalitions of neurons in the cortex and thalamus. The NCC is ‘a minimal set of neural events jointly sufficient for a given conscious experience (given the appropriate enabling conditions)’ (p. 97). A clearly focused perceptual consciousness, for example, is the product of synchronized activity between the inferior temporal and prefrontal cortex. A coalition has the property of reinforcing the firing activity of its member neurons, by synchronizing their spiking discharges and suppressing competing ones. A particular percept stands out because of this ‘winner-takes-all’ competition. Of course, any given percept will have numerous aspects. Each aspect is represented by activity in a particular set of neurons or ‘node’; nodes, however, cannot produce consciousness except as part of a network instantiating a coalition. A conscious percept is therefore associated with a coalition consisting of multifocal activity at numerous nodes, each representing a particular attribute—shape, colour, motion, etc. Thus is achieved the requisite integration of conscious experience without loss of differentiated contents of consciousness.

If you think that you are here being offered an explanation of qualia, think again. Not only does it not even look like such an explanation, Koch does not present it as such. He does not deny the existence of qualia. They are (take a breath) ‘a symbolic form of representation of all of [the] vast ocean of explicit and implicit information associated with the NCC. They come to stand for the penumbra’ (p. 310). Even this rampant anthropomorphism does not take us to the target. While he claims he is seeking to ‘explain how [they] arise from the action of the nervous system’ (p. 304), ‘why qualia feel the way they do remains an enigma’ (p. 310). To me, this seems to be the first, not the last, thing a theory of the relationship between brain activity and subjective experiences should address; after all qualia are what they feel to be.

The KC hypothesis is much more complex than I have described and has been elaborated over many years. There is space here to examine only the ingenious way in which it links the mechanisms by which neural activity qualifies to be conscious with those by which they are unified. Binding is a neural selection mechanism preventing information overload by letting only a fraction of all sensory data pass into awareness. The cellular mechanism of focused attention takes the form of helping one budding coalition establish dominance over other nascent coalitions. Synchrony of activity enables the coalition to achieve dominance: the neurons all sing in harmony. But, as many experimentalists, such as Wolf Singer, have suggested, this may also be the means by which different aspects of a percept are identified as belonging to the same percept and hence are bound together in consciousness. There is some supportive evidence. Within a single visual area, cells coding a continuous stimulus, such as a long sloping line, discharge in synchrony. What is more, cells in different areas coding aspects of a common stimulus, such as the form and movement of a particular object, also resonate...
together. Furthermore, an association between synchronic discharges in thalamo-cortical pathways is correlated with waking consciousness. While the latter is what Crick and Koch might set aside as an ‘enabling condition’, it is interesting that there seems to be an identity of the conditions for binding and for consciousness: that which hold percepts together is what picks them out for figuring in the consciousness of the moment. This avoids the necessity of the relevant fibres dealing with different aspects of the consciousness of, say, an object having to be linked hierarchically through cells that integrate simpler features together into an awareness of a complex object.

Apart from the failure to make sense of qualia, there are other deficiencies in the KC hypothesis. First, as Searle has pointed out (Searle, 1998), even if it turned out to be 100% correct, and consciousness was invariably associated with synchronized neural firing rates (40 Hz seems to be the magic number) in the relevant circuits, this would not feel like an explanation of consciousness. More importantly, the KC hypothesis requires us to imagine all the various impulses firing at the same frequency somehow having this fact available to them in the absence of an observer—an example of ‘the fallacy of misplaced explicitness’ (Tallis, 2004). De facto, simultaneity is not available of itself. Most worrying of all, the assumption of simultaneity as the precursor of the emergence of consciousness assumes that this can be established in the absence of an observer, a difficult position for any physicalist theorist of mind who accepts relativist physics.

There are many other versions of the theory that consciousness arises from an interaction between brain processes (elegantly discussed in Zeman), for example those put forward by Antonio Damasio, Jeffrey Gray and Larry Weiskrantz. Weiskrantz is particularly illuminating. Impressed by the fact that individuals with amnesia or blindsight may exhibit appropriate discriminative behaviour in the absence of explicit recall of learning or relevant visual awareness, he has suggested that what is missing in such subjects is ‘the ability to render a parallel acknowledged commentary on the activities he can still perform’ (quoted in Zeman, p. 290). The homunculus, it appears, is alive and well and talking himself into consciousness. (Edelman, p. 70, describes the dynamic core as a cluster ‘that speaks to itself through an enormous complex of signals across the reentrant network’.)

Given that the seeming explanatory force of most neuroscientific accounts of consciousness relies on wall-to-wall anthropomorphization of parts of the nervous system, where do we go from here? While, as Zeman points out (p. 347), ‘mysterian’ doubts ‘may be barren’, the false fertility reflected in the torrent of writing purporting to explain consciousness is also a kind of barrenness, as Wittgenstein pointed out 50 years ago, when he spoke of ‘experimental methods and conceptual confusion’ of psychology, and of a situation in which ‘problem and method pass one another by’ (Wittgenstein, 1958). This empty fertility will continue until we think again about how we might develop a critical, rather than an uncritical, neurophilosophy.

Time now, perhaps, to switch off the torrent of explanations and step back to consider how we shall reconcile these two large truths: that no account of activity of the brain comes anywhere near an intelligible account of consciousness; and that, nevertheless, of all the objects in the world, the brain seems to be more intimately involved in consciousness—that we are conscious, and what we are conscious of—than any other. If neural activity of a certain kind appears to be a necessary but not a sufficient explanation of consciousness, how shall we investigate the gap between what is necessary and what is sufficient? The investigation should start on the ground floor: with qualia. They are the indubitable stuff of consciousness. Any scientific explanation worth the name should explain ‘the actual feeling of a quale’—‘the warmness of warmth and the greenness of green’ and not wriggle out of this by arguing that scientific stories do not need to provide an explanation any more than physics has to go beyond matter and energy to explain why there is something rather than nothing’ (Edelman, p. 146). What we do not need over the next decade or so are more and more stories with ever more sophisticate homunculi disguised as electronic circuits. And when scientists are tempted, like Crick (echoing Hippocrates), to assert that ‘you, your joys and sorrows, your memories and ambitions, your sense of personal identity and free will, are in fact no more than the behaviour of a vast assembly of nerve cells and their associated molecules’ (Crick, 1994), they should recall that brains are meaningless without bodies; that bodies cannot function without environments; and enviroined human bodies are not human beings without societies. In short, that we shall have to look beyond the brain if we are going to understand what it is to be an aware organism, let alone a self-aware citizen. Which is not to deny, of course, that advances in brain science are helping us better to restore brain-damaged citizens to full citizenship (Academy of Medical Sciences, 2004).

Raymond Tallis
Manchester, UK

DOI: 10.1093/brain/awh311

References


