The Race for Consciousness

by John Taylor, MIT Press, 1999. £19.95 (ix + 360 Pages) ISBN 0 262 20115 1

One thing that makes the scientific problem of consciousness so difficult is that we begin without a fixed point from which to theorize. We want to find an explanation of a complex phenomenon, namely how conscious experience arises in a physical world, yet along the way we must also come to understand that which we hope to explain, consciousness itself. At the outset, that is, we do not know whether the many kinds of conscious experience we have (sensory perceptions, emotional states, memories, dreams, episodes of conscious problem-solving, etc.) constitute a single, unified phenomenon or not. Nor assuming a unified phenomenon, do we know what properties all conscious states share, if any – whether, for example, all conscious states have intrinsic qualitative features or 'raw feels' (and what those might be) or whether all conscious states are transparent and immediate (e.g. when you have the sudden thought 'I forgot to buy milk!' must the content of that thought be self-evident to you – is it known in an unmediated fashion, without inference or introspection?) Finally, given even a 'simple' perceptual experience (e.g. of looking out the window at the rhododendrons in the garden before you), what you experience is not entirely clear. Did you see, in some concrete sense, each individual pink blossom upon the rhododendron or did you merely have an impression of 'lots and lots of pinkish blossoms'[1]? Or is it possible that, as a matter of fact, you did not actually see the blossoms or their colour at all[2–4]? We juggle, that is, two problems at once – both what we hope to explain and our explanation of 'it' – and we pray that an incorrect answer to one problem does not lead us astray in our answer to the other. This is not an unusual conundrum in the neurosciences (e.g. in vision research, we learn both what goes on in vision and how a particular visual system manages to do it) but it does make the task difficult.

John Taylor's book *The Race for Consciousness*, is not, as the title would suggest, a work of science journalism about our most recent scientific infatuation, consciousness. Apart from a brief nod toward the current leading theories, this is Taylor's own entry to the race, his theory of 'relational consciousness'. Taylor's central thesis is deceptively simple: Consciousness arises solely from a process of 'content determination' in which present neural inputs both evoke and intermingle with 'suitable' memories. Filling out just how our multi-faceted experience could arise out of variations on this relational process is the task of the book.

Taylor's strategy is to divide conscious experience into three phenomenal types and their associated relational processes. First, there is 'passive' or perceptual experience, such as tasting your morning coffee or idly looking at the sunset before you. Such experiences are the raw data or phenomenal elements of our conscious experience – the purely qualitative, ineffable and transparent states produced in the posterior cortex, according to Taylor. Second, we have 'self' experience, such as the sudden realization that you ought to drink less coffee and admire more sunsets – in other words, the conscious reflection upon oneself as a subject of experience. Third, there is 'active' consciousness in which the brain actively solves a perceptual or cognitive problem while inhibiting 'self' experience. When we watch a soccer match, for example, the brain must, among other things, decide how to parse the visual scene, selectively attend to the appropriate part (namely what is happening to the soccer ball), figure out just what play is being made, and contain your emotional reactions when your team gives a less than stunning performance.

Taylor's explanation of all three forms of consciousness is roughly the same: a two-step process, followed by a competition mediated by the NRT. For example, in the first stage of passive or perceptual consciousness, pre-processing modules build possible interpretations of an input within a given code, a locally competitive, Hebbian (unconscious) process based upon the synaptic strengths of the neural network at the time of input. Take, for example, a binocular rivalry experiment –one, say, in which the left eye is presented with a pair of parallel lines moving to the left while the right eye is presented with the same stimuli, only moving to the right. Thus, in the pre-processing module for motion (which Taylor hypotheses is visual area MT) both interpretations of the motion input, one of rightward motion and one of leftward motion, are encoded. In the second stage of passive consciousness, rival interpretations are passed on to a working memory buffer where a dominant interpretation emerges through a competitive process. Here, inhibition is the result of contradictory interpretations (e.g. leftward versus rightward motion) and of recent input in the same object code (e.g. a recent motion onset signal). At this juncture, there is a global competition, mediated by the NRT, between the winning interpretations of different working memories. This is where episodic memories from anterior cortex come into play to help select among the global competitors. Consciousness occurs when activity within a working memory reaches a certain level, at the expense of the activity of its global competitors.

The two-step processes Taylor puts forward to explain active consciousness and self consciousness are somewhat less clear to us but perhaps the following will suffice. In active consciousness, working memory is sustained by a series of five cortical thalamic feedback loops in the frontal lobes each of which, Taylor suggests, can be associated with different known functions of the frontal cortex, such the learning and planning actions. First, the contents of passive consciousness – that is, of the contents of posterior working memory – are fed forward to these sites, and then these contents are 'infused' with an action-based semantics, with emotional content from the limbic system, and with the content of relevant episodic memories. (According to Taylor this accounts for the 'greater richness' of the anterior active conscious states than their posterior counterparts.) Consciousness of self, on the other hand, begins with the construction of an autobiographical record, with episodic memories that are strung together in anterior cortex, a record that in effect comes to form one's personality. Our sense of self derives from the active comparison of our present states/responses with those predicted on the basis of our constructed personalities. Again, in both active and passive consciousness, competition between the appropriate working memories decides which contents will become conscious, hence the contribution of each kind of consciousness to the subject's experience at a particular moment in time.

As one might guess from the above synopsis, Taylor's book suffers from a number of problems, not the least of which is expositional clarity. Here, though, we will return to the central problem raised above, of theorizing without a fixed point. For Taylor, the interrelated questions are these: does the relational theory explain consciousness in any deep sense, and is Taylor's grasp on the phenomenon itself, the nature of consciousness, substantive enough to provide a testable theory?

Much of the current debate about consciousness is driven by a distinction proposed by David Chalmers[5] between the 'hard' and 'easy' problems of consciousness. The 'easy' problem is to understand the nature of the neural

events without which there would not be conscious experience, a set of questions with which most neuroscientists are well acquainted, for example: How does the brain bind together into a unified perception the output of various parallel processes occurring in the brain? How is memory stored and retrieved? What are the mechanisms involved in and the function of selective attention? In contrast, the hard question asks for a general explanation of consciousness. *How* is it possible that there are conscious beings in a physical world at all? *Why* do some neural states result in conscious experience and not others? Note that insofar as one believes that there is a hard problem of consciousness, pointing towards the kinds of neural states that correlate with certain kinds of conscious events is not an adequate response. Providing such a correlation is a far from trivial task but, nonethless, the 'how' and 'why' questions would remain open – or so claim those, like Taylor, who grant the distinction.

Taylor's relational model does little to solve the hard problem. He accepts, first, that to explain consciousness '[s]omething more is necessary beyond the mere firing patterns of serried ranks of nerve cells' (p. 122). Then he points to the relations between inputs and memories. Such relations are 'not themselves brain activities' and as a result 'this non-physical essence… opens up the possibility of incorporating a character of seeming insubstantiality into brain activities', that is, of infusing brain activity with contentful consciousness. Exactly why Taylor claims that such relations are nonphysical is unclear, given that the activation of memories by input would seem to be a prototypically causal/physical process. Indeed, even Taylor claims that these relations are 'encoded by the strengths of synapses between neurons activated by related firings', and presumably, the strength of a synapse is also a physical property. But even granting this point, his explanation is wanting. Genuinely 'non-physical' relations are thick on the ground (e.g. Hilary is married to Bill, and she is both luckier than and currently more popular than him) but such relations generally fail to confer 'insubstantiality' on their relata. So the question remains: why would these relations give rise to consciousness?

This leads to the second problem, about the nature of the beast. At the core of Taylor's model is the supposition that, arcane worries aside, consciousness is something with which we are all well acquainted. But if this is not true – if conscious experience is not transparent to the subject – Taylor's theory will run into many problems. Take for example, Taylor's central thesis that memory, in its many guises, contributes to the 'character' of conscious thought. Suppose you hear a familiar tune, 'Twinkle, Twinkle, Little Star'. As you do so, you hear any given note not merely in and of itself, say, as middle C, but as occupying a place within that melody, relative to both the notes already heard and those not yet experienced. Doing so must require memories of many types. But what memories exactly and how do those memories contribute to your auditory experience (as opposed to into the ability to recognize the tune)? Thankfully, we do not hear in our imaginations a cacophony of each and every previous rendition of 'Twinkle, Twinkle Little Star'; nor do we hear a 'generic' rendition, whatever that would be, 'overlaid' upon the present particular one. What we hear is as much a problem for auditory research as how we manage to do so: the contents of our auditory experiences do not seem to be self-evident to us. But without the contents of our experiences in hand, we cannot judge whether the processes of content determination proposed by Taylor really do underlie those experiences. In other words, without transparent access to, or a quantifiable model of, conscious experience itself, Taylor's theory is neither substantive nor testable.

As a final example, consider Taylor's tri-partite division of the consciousness. Take the last of these divisions, our consciousness of self. It is one thing to provide a theory of the neural mechanisms of self, a theory of how and why we manage to both conceive of ourselves and behave as a unified subject, one with a past, present, present and future, with a distinct personality as well as plans and goals. It is another thing to provide a theory of consciousness of the self, of the experience of oneself as such an entity. As Hume rightly pointed out, and as Taylor acknowledges, you do not experience yourself as an introspected object. But that does not preclude experiences of other neural states essential to the maintenance of self; for instance, our ordinary experience of our silent musings and thoughts as our own, as opposed to the pathological experiences of 'hearing voices' or of 'thought insertion'. Which, if any, of our conscious experiences might legitimately be called 'self experiences' is an open and interesting question, not to be decided by an *a priori* tri-partite division of consciousness. Again, without a clear grip on 'that which is to be explained', it is hard to say whether or not Taylor has identified the mechanisms which underlie conscious experience.

Michal Arciszewski and Kathleen Akins
Dept of Philosophy, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6.

References

1 Dennett, D.C. (1991) *Consciousness Explained*, Little, Brown & Co.
2 Grimes, J. (1996) On the failure to detect changes in scenes across saccades. In *Vancouver Studies in Cognitive Science* (Vol. 5) (Akins, K., ed.), Oxford University Press
3 Rensink, R.A. *et al*. (2000) On the failure to detect changes in scenes across brief interruptions. *Visual Cognit.* 7, 127–146
4 O'Regan, J.K. *et al*. (1999) Blindness to scene changes caused by 'mudsplashes'. *Nature* 398, 34
5 Chalmers, D.J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press

Erratum

In the Book Review by Michael Rutter, Vol. 23, pp. 505–506 the price of the book *Neurobiology of Mental Illness* (edited by Dennis S. Charney, Eric J. Nestler and Benjamin S. Bunney) was listed incorrectly as £150; the correct price is £110 or $150. We apologize to the editors and readers for this error.