nature neuroscience

Mysterianism lite

A philosophical view known as 'mysterianism' holds that even though there is nothing supernatural about how consciousness arises from neural activity, the human brain is simply not equipped to understand it. The reason we find the mind–brain problem so baffling, the argument goes, is that humans did not evolve sufficient cognitive abilities to solve it, just as armadillos did not evolve the ability to understand arithmetic. This argument has been advocated by philosophers such as Colin McGinn and cognitive scientists such as Steven Pinker. Now it has been taken up by a prominent science journalist, John Horgan, whose new book *The Undiscovered Mind* offers a view of brain science that might best be described as 'mysterianism lite'. It is not just consciousness that is beyond our grasp, he says; neuroscience as a whole is failing, because the brain is too complicated for human understanding.

Horgan attracted attention, even notoriety, for his 1996 book *The End of Science*, in which he argued that the age of great scientific discoveries is coming to an end because most of the big questions have been answered. The brain is an obvious exception, but Horgan now argues that neuroscience too is reaching its limits, not because it has succeeded in its aims but because those aims are unachievable. The subtitle of his new book is "How the human brain defies replication, medication, and explanation"; its thesis is that the achievements of neuroscience (along with psychology, psychiatry and other related areas) are being oversold, that the supposed practical benefits have been exaggerated, and that the field is now confronting an 'explanatory gap' that may never be bridged. Unlike particle physicists or molecular biologists, says Horgan, neuroscientists "…have yet to achieve their reductionist epiphany. Instead of finding a great unifying insight, they just keep uncovering more and more complexity. Neuroscience's progress is really a kind of anti-progress. As researchers learn more about the brain, it becomes increasingly difficult to imagine how all the disparate data can be organized into a cohesive, coherent whole."

It is tempting to dismiss this as another example of what Richard Dawkins once called "the argument from personal incredulity", but Horgan is surely not alone in finding neuroscience difficult to approach. The brain is immensely complicated, and in the absence of a grand unifying theory for how it works, researchers tend to study very diverse problems that often seem unconnected to each other. It is therefore understandable that their achievements do not always seem intellectually satisfying to nonspecialists.

Part of Horgan's critique may reflect how neuroscience is reported in the media. Among the stories that attract the most attention are the identification of genes or brain areas that are associated with particular behaviors (think of fosB, 'the gene for maternal behavior', or the orbitofrontal cortex, 'the brain's moral compass'), but typically such findings are only the first steps on a long road toward mechanistic understanding. In contrast, many of the most important mechanistic insights into how the brain works (at all levels, from biochemistry to computation) tend to go unreported, because they are very difficult to explain to lay people.

The problem goes deeper than this, though. Even where mechanistic explanations of brain function have been possible, they do not 'feel' like explanations of mental processes. Consider the paper by Treue *et al.* on page 270 of this issue, which presents a striking example of how far our understanding of perception has progressed. Based on knowledge of how motion is represented by populations of neurons in the visual cortex, the authors were able to predict an entirely unexpected visual illusion; two different patterns of moving dots are perceptually indistinguishable, apparently because they both evoke the same pattern of activity in a cortical area called MT. Of course a graph showing the distribution of neuronal firing rates in MT doesn't 'feel' like an explanation of perception. But why should it? The criterion for a good theory is not that it feels right, but that it can successfully predict unexpected results. If a physical theory of neural processing can predict an unexpected mental phenomenon, that is surely a substantial achievement.

It goes without saying that Treue's study raises many further issues—how is the population activity decoded, what other areas are involved in representing the stimuli, and so forth—but there is no reason why questions of this type should not eventually be answered. Certainly, it will be a challenge to understand how (say) a moving red bar is perceived as a unitary stimulus if its orientation, motion and color are each represented in different cortical areas. Horgan may well be right that existing hypotheses to solve this so-called binding problem (such as synchronous oscillations) will prove incorrect. But to deny the possibility of further progress seems perverse. A deeper understanding of the mechanisms underlying mental processes should follow from greater knowledge of anatomical and functional connectivity, better methods of recording and manipulating neural activity, and more realistic computational models, all of which should be achievable with enough time and effort. For ethical reasons, we may never know as much as we would like about human brain activity, but it seems reasonable to expect that insights from other organisms will some day provide good models for much of what happens inside our own heads.

The mysterians might turn out to be correct in claiming that we will never fully understand how brain activity leads to subjective experience: why the firing of MT neurons feels like visual motion, why dopamine release in the nucleus accumbens feels pleasurable, or why electrical stimulation of the anterior supplementary motor area feels amusing. But that is very different from claiming that these phenomena can never be explained in physical terms, or that neuroscience is, as Horgan puts it, "bumping up against fundamental limits of science." Most neuroscientists, fortunately, take a more optimistic view than Horgan; the explanatory gap may not be closed at a single stroke, but it is getting narrower by the day.