ψυχη
PSYCHE

A Review of Jeffrey Gray's *Consciousness: Creeping up on the Hard Problem*

Stephen Biggs
Department of Philosophy
University of Arizona
Social Science Bldg. Rm 213
PO Box 210027
Tucson, Arizona 85721-0027

biggs@email.arizona.edu

PSYCHE 11 (2), March 2005

REVIEW OF: Gray, J. 2004. *Consciousness: Creeping up on the Hard Problem*. Oxford: Oxford University Press. 341 pp. USD$55 (hbk) ISBN: 0198520905.

Jeffrey Gray's *Consciousness: Creeping up on the Hard Problem* will be enjoyed by everyone interested in consciousness. Gray, a neuropsychologist, eloquently summarizes significant experimental results on consciousness and, more importantly, explains both how these results interrelate and how they constrain potential theories of consciousness. He also uses these results to build a novel, fascinating theory of what consciousness does and does not do. Throughout the work Gray's accessible presentation remains deeply respectful of psychologists, neuroscientists, and philosophers' approaches to consciousness. In this respect, Gray's book is an ideal work for an interdisciplinary audience. Sadly, Gray died three months before the publication of this excellent work.

1. A Function for Consciousness

Gray begins *Consciousness: Creeping up on the Hard Problem* by acknowledging Chalmers' (1996) distinction between the hard and easy problems of consciousness. Solving the easy problems requires explaining how we execute certain functions; e.g. how we extract information from sensory stimuli, how we integrate that information, and how we report our thoughts. Solving the Hard Problem requires explaining why there is something it is like for us to execute those functions (or to instantiate specific neural states). Gray proceeds by suggesting that his work creeps up on the Hard Problem's

"scientific" version, which can be solved by explaining how the brain creates qualia, which are the felt properties of experience. (Section 2. below considers the relation between Gray's scientific Hard Problem and Chalmers' traditional Hard Problem.)

Gray creeps up on this problem indirectly both by providing a framework for thinking about qualia in relation to the brain and by asking and answering several related but more tractable questions of qualia, such as "What do they *do*? How did they *evolve*? What *survival value* do they confer?" (p. 67) Since his attempt to answer the first of these questions captures the most interesting and novel aspects of Gray's theory, it is worth considering in detail.

Gray's account of what qualia do begins with an account of what they do not do. Intuition strongly suggests that qualia affect online behavior; seeing a tennis ball causes one to swing one's racket; burning sensations cause one to retract one's hand from the flame; feelings of thirst cause one to drink. Some philosophers find this intuition so compelling that they reject any theory that denies it (e.g. Davidson, 1970). Gray claims, however, that qualia do not affect online behavior. He cites three results (which are more controversial than he acknowledges) to support this counter-intuitive claim.

First, qualia arise too slowly to affect rapid responses. Creating a visual percept, for example, takes time—about 250 milliseconds from retinal stimulation to conscious perception. Performing many online activities, such as returning a fast serve in tennis, requires initiating a motor program more quickly than 250 milliseconds after our sensory apparatus receives relevant stimuli. So, performing such activities requires initiating the motor program *before* consciousness occurs. One might protest that such fast-paced activities—where milliseconds matter—are atypical, and thus, this example is insignificant. Supposing that the facts are correct, however, Gray's example minimally reveals that intuitions about the causal efficacy of qualia sometimes mislead us; for, we intuit that our conscious percept affects our online racket swing, but it does not.

Second, we are wired to behave without qualia. The retina sends information to the brain along several distinct channels. Following Milner and Goodale (1995), Gray thinks of these channels as constituting distinct visual systems. Importantly for Gray, the system that leads to online behavior (the dorsal stream) is distinct from the system that leads to visual qualia (the ventral stream). Accordingly, visual qualia plausibly play no role in affecting online behavior, which is caused, instead, by the dorsal stream. Gray supposes that all senses resemble vision in this regard, and thus, concludes that no qualia affect online behavior.

Third, we become aware of willing an action only after the unconscious brain causes it. Libet (1986) famously found that subjects show a readiness potential for a 'willed' behavior *before* they report becoming aware of willing that behavior. This suggests that the unconscious brain causes behavior independently of our conscious sensation of willing.

How does the unconscious brain accomplish so much? Gray thinks that most behavior is caused by servomechanisms. A thermostat is a simple servomechanism. A thermostat represents both actual temperature and desired temperature (forgive the personification). When the two diverge, it signals the heater. The thermostat subsequently receives new stimulus from the environment and changes its representation

of actual temperature accordingly. When the actual and desired temperatures converge, the thermostat no longer signals the heater. Our brains, thinks Gray, include a wide array of servomechanisms that represent our actual and desired environments, and produce most of our behavior as a result of divergences. It is these mechanisms that lead us to return fast serves, pull our hands from flames, and drink when parched.

Given that unconscious servomechanisms account for online behavior, what do *qualia* do? Gray thinks that qualia allow the unconscious brain to detect errors in servomechanisms' functioning. What sort of error is detected? A thermostat could be designed with an error detector. The detector would expect the thermostat to read a higher temperature after it signals the heater. If the thermostat's assessment of the temperature does not change, the detector concludes that error has occurred. For ease of presentation, suppose that the error results from the thermostat misreading the new environment. The error detector would adjust the thermostat's mechanism for representing the environment accordingly. Because the error detector detects and corrects error only after the thermostat's initial temperature reading and after the heater is signaled and reacts, the error detection is 'late'.

Unless our servomechanisms are infallible, in order for us to function effectively, we must be equipped with a similar *late error detector*. Something in us must be able to recognize when servomechanisms function poorly and alter them accordingly, by changing the servomechanism's assessment of what environmental conditions a stimulus indicates, of what environmental conditions are desirable, or of what output (to, for example, motor programs) produces convergence. Something in us also must be able to reject a servomechanism's reading of the environment when that reading conflicts with readings from other servomechanisms. Performing either task requires a mechanism for comparing expected stimulus to actual stimulus (both within and across servomechanisms).

One might expect Gray to proceed by claiming that consciousness provides the mechanism for late error detection. He claims, instead, that consciousness provides a *medium* in which the unconscious brain can perform such detection. He explains this claim with an analogy,

> Suppose that I am in St. Marks Square in Venice and have sufficient artistic talent . . . to make a passable sketch of it. Later, I use the sketch as an aid to recall St. Marks. Thus the sketch expands my capacity to remember—a causal effect. But the sketch clearly doesn't have this causal effect in its own right. The sketch is made by the brain . . . and later used by the brain. All the causal *mechanisms* lie in the brain. Still, any full description of the causal chain that leads to my recall of St. Marks must include an account of the role played by the sketch. (p. 109)

Just as Gray uses his sketch as a memory aid, the unconscious brain uses consciousness as a display of the recent past. Just as Gray's sketch does not cause him to remember, consciousness does not cause the unconscious brain to adjust servomechanisms. But just as Gray's sketch plays a role in his remembering, so too consciousness plays a role in allowing the unconscious brain to adjust servomechanisms. Specifically, consciousness

allows the unconscious brain to recognize that error has occurred such that it can adjust servomechanisms effectively.

What makes consciousness the right candidate for this task? Gray thinks that consciousness possesses several relevant features, including the following two. First, whereas servomechanisms track rapid changes in the environment, qualia model the environment's relatively enduring and stable features—which explains why consciousness occurs so slowly: building such an abstract model takes time. Accordingly, consciousness preserves (and replays) the past and thereby creates a medium for comparison of expectations with actualities; by creating a record of past stimuli and behaviors, consciousness creates a medium in which the unconscious brain can compare servomechanisms' current inputs with the inputs that would be expected given their past inputs and outputs. Second, whereas servomechanisms are isolated from one another, consciousness integrates information from all servomechanisms, and thus, the conclusions of various servomechanisms also can be compared.

Gray's argument, then, is rather simple. Given our fallibility, we must have a mechanism for late error detection. This mechanism can function properly only with the aid of a model of the enduring features of the world that *inter alia* integrates information from various sensory modalities. Consciousness provides this model. So, consciousness functions as a display medium that allows the unconscious brain to perform late error detection by comparing expectations to actualities.

2. Late Error Detection and the Hard Problem

This account of qualia *qua* medium for late error detection could solve (or at least address) the traditional Hard Problem in either of two ways, through an appeal to selective fitness or through an appeal to philosophical functionalism. First, given two suppositions about consciousness and selective fitness, this account explains why consciousness exists. Suppose first that we would not survive if we could not perform late error detection. Suppose next that *only* consciousness could possess the features required of the model. It follows that consciousness is a necessary condition for survival, and thus, its existence should not be surprising. Gray, however, denies the second supposition as implausibly strong.

Second, this account solves the traditional Hard Problem if qualia are identical with functions—for, by explaining why the functions exists and which functions constitute qualia, Gray would thereby explain why qualia exist. Gray, however, rejects the claim that qualia are identical with functions. Specifically, he rejects the implication that each type of conscious state is identical with a type of functional state. Synaesthetes, he notes, can instantiate the same conscious state type without instantiating the same functional state type. In certain synaesthetes, a yellow quale can be caused either by seeing a ripe banana or by hearing a middle C. Depending on which stimulus causes the quale, moreover, the synaesthete reacts to it in quite different ways. Accordingly, the synaesthete can instantiate a single conscious state type without instantiating the same functional state type. Thus, conscious states are not identical with functional states. (Of course, a functionalist could deny the claim that synaesthetes experience the same qualia in the two functional states, or they could adopt a supervenient form of functionalism.)

Since Gray rejects the claim that only consciousness could provide a model for late error detection, his recognizing that consciousness provides this model does not explain why consciousness exists—at least, not for him. Since Gray rejects the claim that conscious states are identical with functional states, his finding the proper function of consciousness does not explain why consciousness exists—again, at least, not for him. But Gray does little more *vis-à-vis* the Hard Problem than identify this function of consciousness. So, although Gray situates his theory against the background of the Hard Problem, it does not solve the Hard Problem.

Gray, in fact, accepts that his work does not *solve* the Hard Problem. He insists, however, that recognizing the function of consciousness constitutes *creeping up on* the Hard Problem. One can think about Gray's insistence in either of two ways. First, the less generous reader will conclude that Gray has confused the traditional Hard Problem with easy problems. Second, the more generous reader will search for a way in which Gray's theory helps us creep on the traditional Hard Problem. Although relevant obscurities are the weakest aspect of Gray's book, I favor the generous reading.

There are, however, two *prima facie* compelling reasons to think that Gray has confused the Hard Problem with easy problems. First, a literal interpretation of Gray's formulation of the Hard Problem, which asks '*how does the brain create qualia?*' (p. 301), invites primarily easy problems. Second, while identifying the function of consciousness clearly addresses easy problems, it is not clear that it helps to explain why qualia exist. Both of these claims must be dismissed before endorsing the generous reading.

According to Gray, explaining how the brain creates qualia solves the scientific version of the Hard Problem. If Gray is right to call this 'a version of the Hard Problem', then explaining *how* the brain creates qualia also explains (or at least addresses) *why* there is something it is like to execute certain functions (or to be in certain neural states) or *why* specific functions (or neural states) produce specific qualia. But this consequent is *prima facie* implausible—explaining how rarely explains why, answering questions about mechanism rarely satisfies philosophical concerns. Gray, moreover, never explicitly considers the relation between his version of the Hard Problem and Chalmers', except to claim that the former *is* a version of the latter.

Several aspects of Gray's work suggest, nevertheless, that, for Gray, explaining *how* the brain creates qualia requires explaining *why* the brain creates qualia. For example, unlike many neuropsychologists, he does not think that finding the neural correlates of consciousness (thereby) explains how the brain creates qualia, even given the supposition that these neural correlates cause qualia. What more would explaining how the brain creates qualia require? Gray seems to think that it requires explaining both *why* these neural states, as opposed to other neural states or automobiles, produce qualia, and *why* particular neural states produce particular qualia. These explanations clearly would address the traditional Hard Problem.

Gray's dedicated review of Stuart Hameroff and Roger Penrose's quantum-consciousness theory supports this assessment. Although Gray concludes that their theory is likely inaccurate, he thinks that it is the only current scientific theory with the resources required to explain how the brain creates qualia. What resources does this

theory possess uniquely? To oversimplify, by placing (proto-) qualia at the fundamental level of the universe, this theory can explain how qualia appear at the top. Philosophers (e.g. Russell, 1927) have long recognized this as a way to explain not merely how the brain creates qualia but also *why* the brain creates qualia.

If, however, we accept that Gray intends for his scientific Hard Problem to resemble Chalmers' traditional Hard Problem, the second concern arises; that is, while identifying the function of consciousness clearly addresses easy problems, it is not clear that it helps to explain *why* qualia exist. How might identifying the function of consciousness address the traditional Hard Problem? Gray could have argued that, even though other aspects of our brain could have provided a medium for the unconscious brain to perform late error detection, they did not, and thus, qualia exist because they perform this function. Alternatively, he could have argued that identifying the function of conscious provides a necessary background for scientific investigations of qualia: for, investigations of the Hard Problem include study of the phenomena that create it, qualia; plausibly, scientists can investigate qualia most effectively *indirectly*, by studying its functional (or neural) correlates; so, scientific investigation of the Hard Problem presupposes identification of the function of consciousness. Gray, however, does not pursue either possibility, both of which would require significant development. Instead, Gray simply insists that finding the function of and consequent survival value for qualia constitutes creeping up on the Hard Problem. Although many readers (including me) will share this intuition, Gray's work would have benefited from a thorough exploration of this claim.

In the end, readers must decide for themselves both how Gray's scientific Hard Problem relates to its traditional counterpart and which aspects of Gray's work address which of the many problems of consciousness. Whatever one decides, she will enjoy and benefit from reading *Consciousness: Creeping up on the Hard Problem*.

References

Chalmers, D. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Davidson, D. 1970. Mental Events. In D. Rosenthal (Ed.) *The Nature of Mind*. (p. 247-256). New York: Oxford University Press.

Milner, D. & Goodale, M. 1995. *The Visual Brain in Action*. Oxford: Oxford University Press.

Russell, B. 1927. *An Analysis of Matter*. New York: Routledge.

In his book 'Consciousness: Creeping up on the Hard Problem' written towards the end of his life, Gray summarised his ideas about brain function and consciousness. He took the view that the contents of consciousness are usually about something, and this is described as intentionality or meaning. He suggested that intentionality is another aspect of the 'binding problem', as to how the different modalities, such as sight and hearing, are bound together into a single conscious experience. Gray argued that without such binding, eating a banana could involve seeing yellow, feel Consciousness is suggested to be the component of this field that is transmitted back to neurons, and communicates its state externally. Thoughts are viewed as electromagnetic representations of neuronal information, and the experience of free will in our choice of actions is argued to be our subjective experience of the cemi field acting on our neurons.Â His article underwent peer review before publication. In fact, Baars is on the editorial board of the journal that published it. The field theories of consciousness do not appear to have been as widely discussed as other quantum consciousness theories, such as those of Penrose, Stapp or Bohm.[19][20][21] However, David Chalmers[22] argues that quantum theories of consciousness suffer from the same weakness as more conventional theories.