Draft: Comments welcome
Please do not quote

Conscious Realism and the Mind-Body Problem

Donald D. Hoffman

Department of Cognitive Sciences

University of California

Irvine, California 92697

ddhoff@uci.edu

Phone: 949-824-6795

Fax: 949-824-2307

Abstract
Despite substantial efforts by many researchers, we still have no scientific theory of how brain activity can create, or be, conscious experience. This is troubling, since we have a large body of correlations between brain activity and consciousness, correlations normally assumed to entail that brain activity creates conscious experience. Here I explore a solution to the mind-body problem that starts with the converse assumption: these correlations arise because consciousness creates brain activity, and indeed creates all objects and properties of the physical world. To this end, I develop two theses. The *multimodal user interface (MUI*) theory of perception states that perceptual experiences do not match or approximate properties of the objective world, but instead provide a simplified, species-specific, user interface to that world. *Conscious realism* states that the objective world consists of conscious agents and their experiences; these can be mathematically modeled and empirically explored in the normal scientific manner.

Keywords: consciousness, conscious realism, epiphenomenalism, epiphysicalism, evolution, natural selection, neural correlates of consciousness, perception, physicalism, user interface

Introduction

What is the relationship between consciousness and biology? This question, a version of the classic mind-body problem, has in some form troubled philosophers at least since the time of Plato, and now troubles scientists. Indeed, a list of the top 125 open questions in *Science* puts the mind-body problem at number two, just behind the question, "What is the universe made of?" (Miller, 2005). The mind-body problem, as *Science* formulates it, is the question, "What is the biological basis of consciousness?"

The reason for this formulation is, in part, the large and growing body of empirical correlations that have been found between consciousness and brain activity. For instance, it has been found that damage to cortical area V1 is correlated with loss of conscious visual perception (Celesia et al., 1991). If V1 is intact but certain extrastriate cortical regions are damaged, there is again loss of conscious visual perception (Horton & Hoyt, 1991). Damage to the lingual and fusiform gyri are correlated with achromatopsia, a loss of color sensation (Collins, 1925; Critchley, 1965), and magnetic stimulation of these areas is correlated with chromatophenes, conscious experiences of unusual colors (Sacks, 1995, p. 28; Zeki, 1993, p. 279). Damage to area V5 is correlated with akinetopsia, a loss of motion sensation (Zihl et al., 1983, 1991; Rizzo et al., 1995); magnetic inhibition of V5 is also correlated with akinetopsia (Zeki et al., 1991). In many tasks in which subjects view displays with binocular rivalry, so that they consciously perceive the stimulus presented to one eye and then periodically switch to consciously perceive the stimulus presented to the other eye, there are changes in cortical activity precisely correlated with changes in conscious perception (Alais & Blake, 2004), changes that can be measured with fMRI (Lumer et al., 1998; Tong et al.,, 1998), EEG (Brown & Norcia, 1997), MEG (Tononi et al., 1998), and single unit recording (Leopold & Logothetis, 1996); such correlated activity can be found in ventral extrastriate, parietal, and prefrontal cortices (Rees et al., 2002). Activity in brain systems with a high degree of information integration is correlated with conscious experience (Tononi & Sporns, 2003), including, most notably, reentrant connections between posterior thalamocortical systems for perceptual categorization and anterior systems for categorical memory (Edelman, 1987), whereas activity in systems with a low degree of such integration is not correlated with conscious experience (Tononi & Sporns, 2003).

Such correlations, and many more not mentioned here, persuade most researchers that brain activity causes, or is somehow the basis for, consciousness. As Edelman (2004, p. 5) puts it, "There is now a vast amount of empirical evidence to support the idea that consciousness emerges from the organization and operation of the brain… The question then becomes: What features of the body and brain are necessary and sufficient for consciousness to appear?" Similarly, Koch (2004, pp. 1–2) argues, "The fundamental question at the heart of the mind-body problem is, *what is the relation between the conscious mind and the electro-chemical interactions in the body that give rise to it?* How do [conscious experiences] emerge from networks of neurons?"

Consensus on this point shapes the current scientific statement of the mind-body problem. It is not the neutral statement that opened this section, viz., "What is the relationship between consciousness and biology?" Instead, as *Science* makes clear, it is a statement that indicates the expected nature of the solution, "What is the biological basis of consciousness?"

Given this consensus, one would expect that there are promising theories about the biological basis of consciousness, and that research is proceeding to cull and refine them. Indeed such theories are numerous, both philosophical and scientific, and the volume of empirical work, briefly highlighted above, is large and growing.

For instance, following the demise of behaviorism in the 1950s, there have been several classes of philosophical theories of the mind-body problem. Type physicalist theories assert that mental state types are numerically identical to certain neural state types (Place, 1956; Smart, 1959). This identity claim has seemed, to many philosophers, too strong. It seems premature to dismiss the possibility that creatures without neurons might have mental states, or that the same mental state type might be instantiated by different neural state types in different people or animals. Such considerations led to the weaker token physicalist theories, which assert that each mental state token is numerically identical to some neural state token (Fodor, 1974). Reductive functionalist theories assert that the type identity conditions for mental states refer only to relations, typically causal relations, between inputs, outputs, and each other (Block & Fodor, 1972). Nonreductive functionalist theories make the weaker claim that functional relations between inputs, outputs and internal system states give rise to mental states but are not identical with such states (Chalmers, 1995). These theories typically entail epiphenomenalism, the claim that conscious experiences are caused by neural activity but themselves have no causal consequences. This is thought by some to be a reductio of nonreductive functionalism, since it entails that a person's beliefs about their conscious experiences are not caused by those experiences, and indeed their beliefs would be the same even if they had no such experiences. Representationalist theories (e.g., Tye, 1995, 2000) identify conscious experiences with certain tracking relationships, i.e., with certain causal covariations, between brains states and states of the physical world. On these theories it is the entire causal chain, not just the neural activity, that is to be identified with, or gives rise to, conscious experience. The "biological naturalism" theory of Searle (1992, 2004) claims that conscious states are caused by lower level neural processes in the brain. Single neurons are not conscious, but some neural systems are conscious. Consciousness can be causally reduced to neural processes, but it cannot be eliminated and replaced by neural processes.

This brief overview does not, of course, begin to explore these theories, and it omits important positions, such as the emergentism of Broad (1925), the anomalous monism of Davidson (1970), and the supervenience theory of Kim (1993). However it is adequate to make one obvious point. The philosophical theories of the mind-body problem are, as they advertise, philosophical and not scientific. They explore the conceptual possibilities where one might eventually formulate a scientific theory, but they do not themselves formulate scientific theories. The token identity theories, for instance, do not state precisely which neural state tokens are identical to which mental state tokens, together with principled reasons why. The nonreductive functionalist theories do not state precisely which functional relations give rise, say, to the smell of garlic versus the smell of a rose, and do not give principled reasons why, reasons that lead to novel, quantitative predictions. These comments are not, of course, intended as a criticism of these theories, but simply as an observation about their intended scope and limits.

It is from the scientists that we expect theories that go beyond statements of conceptual possibilities, theories that predict, from first principles and with quantitative precision, which neural activities or which functional relations cause which conscious experiences. Scientists have produced several theories of consciousness.

For instance, Crick and Koch (1990; Crick, 1994) proposed that certain 35–75 hertz neural oscillations in cerebral cortex are the biological basis of consciousness. They noted that such oscillations seem to be correlated with conscious awareness in vision and smell, and that they could instantiate a solution to the binding problem, viz., the problem of integrating perceptual information—such as color, motion, or form—that is represented in separate cortical areas, to create unified perceptions of objects. Subsequently Crick and Koch (2005) proposed that the claustrum may be responsible for the unified nature of conscious experience. The claustrum receives inputs from nearly all regions of cortex and sends projections to nearly all regions of cortex, a pattern of connectivity ideal for integrating widespread cortical activity into a unified conscious experience.

A different neural theory, the theory of the dynamic core, has been proposed by Edelman and Tononi (2000). It states that, "A group of neurons can contribute directly to conscious experience only if it is part of a distributed functional cluster that, through reentrant interactions in the thalamocortical system, achieves high integration in hundreds of milliseconds." (p. 144). Furthermore, according to this theory, "To sustain conscious experience, it is essential that this functioning cluster is highly differentiated, as indicated by high values of complexity." (p. 144). They give a mathematical formulation of complexity, a formulation that has since been refined and renamed a measure of information integration (Tononi & Sporns, 2003).

Baars (1988) proposed that consciousness arises from the contents of a global workspace, a sort of blackboard by which various unconscious processors communicate information to the rest of the system. The global accessibility of the contents of this blackboard is the source of conscious experience.

Hameroff and Penrose (1996; Penrose, 1994) proposed that quantum coherence and quantum-gravity-induced collapses of wave functions are essential for consciousness. They suggest that tubulins within neuronal microtubules are coupled to internal quantum events and interact with each other in both classical and quantum fashion. This allows the tubulins to implement noncomputable functions, which Hameroff and Penrose suggest are also essential for consciousness.

Stapp (1993, 1996) proposes that a main task of the alert brain is to construct, at each moment, a template for the next action of the organism. The brain, being itself a quantum system, naturally evolves a superposition of such action templates. The collapse of this superposition of templates to a unique template gives rise to conscious experience.

Again, this brief overview does not begin to explore these theories and, for brevity, omits some. But the pattern that emerges is clear. The theories so far proposed by scientists are, at best, hints about where to look for a genuine scientific theory; none remotely approaches the minimal explanatory power, quantitative precision, and novel predictive power expected from a genuine scientific theory. We would expect, for instance, that a scientific theory of consciousness would be able to explain, at least in principle, the difference in conscious experience between, e.g., the smell of a rose and the taste of garlic. How, precisely, is the smell of a rose generated by a 40 hertz oscillation, a reentrant thalamocortical circuit, a certain level of information integration, an entry in a global workspace, the quantum state of microtubules, or the collapse of evolving templates? What precise changes in these must take place to change the experience from the smell of a rose to the taste of garlic? What quantitative principles account for these changes? We are not here asking about advanced features of consciousness, such as self consciousness, that are perhaps available to just a few species. We are asking about a most elementary feature, a feature we would expect to find even in a rat. But none of the theories proposed by scientists has the tools to answer these questions and none, as yet, even gives us any guidance how to build such tools. They do not begin to dispel the mystery of conscious experience. As Pinker points out, "Sentience is not a combination of brain events or computational states: how a red-sensitive neuron gives rise to the subjective feel of redness is not a whit less mysterious than how the whole brain gives rise to the entire stream of consciousness." (Pinker, 1997, p. 564).

In short, the scientific study of consciousness is in the embarrassing position of having no scientific theory of consciousness. The existing theories are just hints that, for the moment, give us no foreseeable way to answer the most basic questions about conscious experiences. It is not that we have several scientific theories and are looking to the empirical data to cull and refine them. Rather, we cannot yet even formulate a single scientific theory, and we cannot envision how it might be done.

This remarkable situation has led to several responses. The first is to conclude that, although consciousness arises naturalistically from brain activity, humans are not equipped with the cognitive capacities required to formulate an adequate scientific theory. The relation between consciousness and brain activity seems mysterious not because it is in fact supernatural, but because of limits in our cognitive capacities. As McGinn (1989) puts it, "We know that brains are the de facto causal basis of consciousness, but we have, it seems, no understanding whatever of how this can be so." Pinker agrees with this assessment. After asking how conscious experience arises from physical systems he answers, "Beats the heck out of me. I have some prejudices, but no idea of how to begin to look for a defensible answer. And neither does anyone else. The computational theory of mind offers no insight; neither does any finding in neuroscience, once you clear up the usual confusion of sentience with access and self-knowledge." (Pinker, 1997, pp. 146–147). And later Pinker adds, "Our thoroughgoing perplexity about the enigmas of consciousness, self, will, and knowledge may come from a mismatch between the very nature of these problems and the computational apparatus that natural selection has fitted us with." (ibid., pp. 565).

This also seems to have been the view of Thomas Huxley (1866): "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of Djin when Aladdin rubbed his lamp." To solve the mind-body problem, on this view, we do not need to work harder. Instead, we need a mutation that endows us with the necessary concepts. This position cannot, for now, be ruled out, given that we have no scientific theory.

A second response is to conclude that we must keep trying until we find the empirical fact that leads to a theoretical breakthrough. When Francis Crick was writing *The astonishing hypothesis*, which presents his neural theory of consciousness, I asked him over lunch if he could explain how neural activity gives rise to specific conscious experiences, such as the experience of the color red. When he replied no, I asked him if, given the liberty to make up any new neurobiological discovery he wished, he could fabricate a discovery that would allow him to create such an explanation. He again replied no. But he then quickly pointed out that he was not impressed by arguments from poverty of the imagination, and urged that we must press forward with the study of neurobiology until we make the crucial empirical discoveries. This is a defensible position and, indeed, the position of most researchers in the field.

A third response claims there is no mind-body problem, on at least two different grounds: There is no mind to reduce to body, or no body to which mind can be reduced.

The first is sometimes asserted by eliminative materialists, who claim that nothing in reality corresponds to the categories of our folk psychology (P.M. Churchland, 1981; P. S. Churchland, 1986), and in particular nothing corresponds to our folk notions about conscious experiences (Dennett, 1978). As neuroscience progresses we will not find smooth reductions of our notions of conscious experiences to neural activity; we will find that such notions are hopelessly false and must be abandoned, much as we abandoned phlogiston or the claim that the sun rises, and we will instead use only the language of neurophysiology. The plausibility or implausibility of this claim for beliefs and other propositional attitudes, although an interesting topic, is beside the point here. Our issue is the plausibility of eliminating conscious experience, and here there is a fundamental divide between those who find it plausible and those who don't, with little productive discourse between them. I find elimination implausible: if I know anything, I know I have conscious experience. But such experience has also shown me that little is gained by argument with those who disagree.

The second claim, that there is no body to which mind can be reduced, is made most notably by Chomsky (1980, 2000), who argues that there has not been a coherent formulation of the mind-body problem since Newton, who, by introducing action at a distance, destroyed the mechanical philosophy. Since then, there has been no principled demarcation between the physical and nonphysical, and hence no clear notion of body to which mind can be reduced. Mind was unaffected by Newton's theory, but body was destroyed. The natural reaction, according to Chomsky, is to view consciousness as a property of organized matter, no more reducible than rocks or electromagnetism (Chomsky, 2000, p. 86). Critics of this view have replied that it is biology, not physics, that most probably will count as body for the mind-body problem, and further advances in physics are unlikely to alter the aspects of biology relevant to the solution of this problem (Smart, 1978). This reply misses Chomsky's main point, which is that no distinction can now be made between the physical and mental; advances in physics will change this situation only if they lead to a principled distinction, as was once had in the contact physics of Descartes (but see Lycan, 2003). Chomsky's point here is well taken. But given this, it seems inconsistent for Chomsky to further insist that consciousness is a property of organized matter. First, what counts as matter is not much clearer than what counts as physical. Second, even if we could clearly define matter, why should we necessarily expect, in the nondualistic setting that Chomsky endorses, that consciousness should be a property of organized matter rather than vice versa?

This is a natural point of departure for the theory developed here. The dualistic formulation of the mind-body problem, in which conscious experience arises from non-conscious neurobiology or physics, has failed, despite prodigious efforts, to produce a single scientific theory. Nevertheless, the vast majority of researchers continue to pursue a solution from within this dualistic framework. They might turn out to be right. But the search space of possible scientific theories of the mind-body problem is large, and it is reasonable, given the failure, so far, of all explorations in the dualistic region, for a few researchers to explore elsewhere. That is the intent here: to explore a nondualistic, but mathematically rigorous, theory of the mind-body problem, one that does not assume from the outset that consciousness is a property of organized matter. To do this requires us first to develop a nondualistic theory of perception. We begin this development by questioning a key assumption of current perceptual theories.

Perception as faithful depiction

Current scientific theories of perception fall into two main classes: direct and indirect (see, e.g., Fodor & Pylyshyn, 1981; Hoffman, 1998; Palmer, 1999).

Indirect theories, which trace their lineage through Helmholtz (1910/1962) and Alhazen (956–1039/Sabra, 1978), typically claim that a goal of perception is to match, or at least approximate, useful properties of an objective physical world (Marr, 1982). The physical world is taken to be objective in the sense that it, and its properties, do not depend on the perceiver for their existence. According to indirect theories, the data transduced at the various sensory receptors is not sufficiently rich, by itself, to determine a unique and correct match or approximation. Therefore the perceiver must infer properties of the world on the basis of certain constraining assumptions. For instance, in the perception of the 3D shapes of objects from their visual motions, the perceiver might infer shape on the basis of a *rigidity* assumption: If the image data could have arisen, in principle, by projection of the motion of a rigid 3D body, then the visual system infers that the image data are, in fact, the projection of that rigid body (Ullman, 1979). This inference might be couched in the mathematical framework of regularization theory (Poggio, Torre, & Koch, 1985) or Bayesian inference (Knill & Richards, 1996). Such inferences can be quite sophisticated and computationally intensive, which might account, in part, for the fact that more than half of the human cerebral cortex is engaged in perception.

Direct theories, which trace their origin to Gibson (1950, 1966, 1979/1986), agree with indirect theories that a goal of perception is to match an objective physical world, but disagree with the claim that the sensory data are too impoverished, by themselves, to do the job. Instead, direct theorists argue that the sensory data are sufficiently rich to uniquely determine the correct specification of the state of the world, although this specification is primarily for affordances, those aspects of the world most relevant to the needs and actions of the perceiver. Thus the perceiver can, without any inferences, simply pick up true properties of the objective environment directly from the sensory data. What is this environment? "The environment consists of the earth and the sky, with objects on the earth and in the sky, of mountains and clouds, fires and sunsets, pebbles and stars ... the environment is all these various things—places, surfaces, layouts, motions, events, animals, people, and artifacts that structure the light at points of observation." (Gibson, 1979, p. 66).

The debate between these two classes of theories raises interesting issues (Fodor & Pylyshyn, 1981; Ullman, 1980). But what is pertinent here is that both agree that a goal of perception is to match or approximate true properties of an objective physical environment. We can call this the *hypothesis of faithful depiction*. This hypothesis is widespread and rarely questioned now in the scientific study of perception.

For instance, Stoffregen and Bardy (2001) state: "We analyze three hypotheses about relations between ambient arrays and physical reality: (1) that there is an ambiguous relation between ambient energy arrays and physical reality, (2) that there is a unique relation between individual energy arrays and physical reality, and (3) that there is a redundant but unambiguous relation, within or across arrays, between energy arrays and physical reality." The first hypothesis is endorsed by indirect theories, and the second by some direct theories. They conclude in favor of the third hypothesis, viewing it as an extension of standard direct theories. Nowhere do they question the assumption of faithful depiction that is shared by all three; nor do any of the more than 30 commentaries on their article.

Yuille and Buelthoff (1996, p. 123) start their description of the Bayesian approach to perception with the hypothesis of faithful depiction: "We define vision as perceptual inference, the estimation of scene properties from an image or sequence of images … there is insufficient information in the image to uniquely determine the scene. The brain, or any artificial vision system, must make assumptions about the real world. These assumptions must be sufficiently powerful to ensure that vision is well-posed for those properties in the scene that the visual system needs to estimate." The objective physical environment has certain properties, and perception uses Bayesian estimation to recover, or reconstruct, those properties from sensory data. The commitment to the hypothesis of faithful depiction is clear in such terms as 'estimate', 'recover', and 'reconstruct', which appear repeatedly throughout the literature of computational vision.

Lehar (2003) discusses the theoretical gap between consciousness and current models of neurobiology, and proposes a "Gestalt Bubble" perceptual modeling approach to bridge the gap. This model, he concludes, entails the hypothesis of faithful depiction for spatial properties of the visual scene: "The perceptual modeling approach reveals the primary function of perception as that of generating a fully spatial virtual-reality replica of the external world in an internal representation." (p. 375).

Purves and Lotto (2003) appear, on first reading, to reject the hypothesis of faithful depiction. They reject, for instance, "the seemingly sensible idea that the purpose of vision is to perceive the world as it is…" (p. 5). They suggest instead that "what observers actually experience in response to any visual stimulus is its accumulated statistical meaning (i.e., what the stimulus has turned out to signify in the past) rather than the structure of the stimulus in the image plane or its actual source in the present" (p. 287). Thus Purves and Lotto do not, in fact, recommend rejection of the hypothesis of faithful depiction *tout court*. They simply recommend rejecting a version of the hypothesis that focuses exclusively on the *present* stimulus and the *present* state of the physical world. They do endorse a version of the hypothesis that includes an adequate statistical sample of the past. On their version, a goal of perception is to approximate the true statistical properties of an objective physical environment *over an appropriate history*. The hypothesis of faithful depiction is not being challenged, only versions that restrict its time frame to the present moment. The purpose of vision is to perceive the world, not just as it is, but as it has been.

Noë and Regan (2002) also appear, on first reading, to reject the hypothesis of faithful depiction. They reject, for instance, the position that "…the visual system builds up a detailed internal representation of the three-dimensional environment on the basis of successive snapshot-like fixations of the scene …" (p. 575). They propose instead that "what one sees is the aspect of the scene to which one is attending—with which one is currently interacting…" (p. 575). Thus Noë and Regan also do not reject the hypothesis of faithful depiction tout court. They claim that "Perceivers are right to take themselves to have access to environmental detail and to learn that the environment is detailed" (p. 576) and that "the environmental detail is present, lodged, as it is, right there before individuals and that they therefore have access to that detail by the mere movement of their eyes or bodies" (p. 578). Thus they support a version of the hypothesis of faithful depiction that is careful to observe the limits of perceptual attention and the critical role of sensorimotor interactions.

The hypothesis of faithful depiction is so universally accepted that it appears in the textbooks. In his standard textbook, *Vision Science,* Palmer (1999, p. 6) endorses the hypothesis of faithful depiction as follows: “Evolutionarily speaking, visual perception is useful only if it is reasonably accurate … Indeed, vision is useful precisely because it is so accurate. By and large, *what you see is what you get.* When this is true, we have what is called **veridical perception** … perception that is consistent with the actual state of affairs in the environment. This is almost always the case with vision…” [emphases his]. Palmer goes on to explain that vision accomplishes this faithful depiction not by passive mirroring but by active construction.

The hypothesis of faithful depiction is endorsed not just by scientists, but also by some philosophers. Searle (2004, p. 171), for instance, says: “In visual perception, for example, if I see that the cat is on the mat, I see how things really are (and thus achieve mind-to-world direction of fit) only if the cat’s being on the mat causes me to see the situation that way (world-to-mind direction of causation).”

I, too, have endorsed the hypothesis of faithful depiction, describing the central questions about visual perception as follows: “First, why does the visual system need to organize and interpret the images formed on the retinas? Second, how does it remain true to the real world in the process? Third, what rules of inference does it follow?” (Hoffman , 1983, p. 154).

I now think the hypothesis of faithful depiction is false. It is not a goal of our perceptual systems to match or approximate properties of an objective physical world. Moreover evolutionary considerations, properly understood, do not support the hypothesis of faithful depiction, but instead require its rejection.

I propose instead that perception is a multimodal user interface (Hoffman, 1998; 2003). A successful user interface does not, in general, match or approximate what it represents. Instead it dumbs down and reformats in a manner useful to the user. It is because it simplifies, rather than matches, that the user interface usefully and swiftly informs the actions of the user. Moreover the properties employed in the user interface can be, and often are, entirely distinct from those of the represented domain, with no loss of effectiveness. A perceptual user interface, dumbed down and reformatted appropriately for the niche of a particular organism, gives that organism an adaptive advantage over one encumbered with the job of constructing a match or approximation to some aspect of the complex objective world. The race is to the swift; a user interface makes one swift precisely by not matching or approximating the objective world.

This is not what the textbooks, or most perceptual experts, say and therefore invites some spelling out. I begin by discussing user interfaces and virtual worlds.

User interfaces

Suppose you wish to delete a file on your PC. You find the icon for the file, click on it with your mouse, drag it to the recycle-bin icon, and release. Quick and easy. The file icon might be blue and square. The recycle bin might be shaped like a trash can. All for ease of use. Of course what goes on behind the icons is quite complex: A central processor containing millions of transistors reads and executes binary commands encoded as voltages in megabytes of memory, and directs the head on a hard drive that has a disk revolving thousands of times per minute to move to a specific place over the disk and make changes to its magnetic structure. Fortunately, to delete a file you don’t need to know anything about this complexity. You just need to know how to move colorful icons.

The icons, and the entire graphical-windows interface, is designed to help the user by hiding the complexity of the computer (see, e.g., Schneiderman, 1998). This is accomplished, in part, by *friendly formatting.* The windows interface and its contents are designed not to resemble the actual complexity of the computer and its inner workings, but instead to present needed information to the user in a format that is friendly, i.e., that is easy and natural to use. Although the actual file in the computer is a complex array of voltages and magnetic fields with no simple geometry, the file icon is a square because this is a simple conventional symbol easily interpreted by human users. Nothing about the shape of the file icon resembles the shape of the file itself. This is no failure of the icon, no gross misrepresentation of reality. It is, instead, what makes the icon useful. Few souls delight to search the guts of a computer with voltmeter and magnetometer to find a file. We much prefer to seek a square icon in a pretty display. Again, the file icon might be blue. But nothing about the file itself, the voltages and magnetic fields inside the computer, is blue. Is this gross misrepresentation by the icon? Of course not. The color of the icon is not intended to resemble anything about the file but simply to indicate, say, what kind of file it is or how recently it was last modified. The icon sits at some spot on the display, perhaps in the upper right. But this does not mean that the file itself is in the upper right of the computer. The location of an icon on the display is, in part, simply a convenient way to keep track of it. There is, in short, no resemblance between properties of the icon and properties of the file. This is no problem, no failure of veridicality. It is the intended consequence of friendly formatting.

The interface also helps the user by means of *concealed causality*. Not only is the structural complexity of the computer hidden behind the icons, but also its causal complexity. When you drag the file icon to the recycle bin and release, does the movement of the file icon to the recycle bin icon cause the deletion of the file? No. The icons have no causal powers within the computer. They are just patterns of pixels on the display, and send no signals back to the computer. The complex causal chain within the computer that ultimately deletes the file is completely hidden, behind the interface, from the user. And nothing in the movement of the file icon to the recycle-bin icon remotely resembles anything in this complex causal chain. Is this a failure or misrepresentation of the interface? No. To the contrary, it is the reason for the interface. Hiding the true causal complexity helps the user to quickly and easily delete a file, create a new one, modify an illustration, format a disk, and other such actions without being slowed and distracted by a myriad of causal details.

Although the icons of the interface have no causal powers they are nonetheless useful by providing *clued conduct*. The icons effectively inform actions of the user, allowing the user to trigger the appropriate, but hidden, causal chains. In the case of deleting a file, the icon of the file informs the user how to click the mouse, and the icon of the recycle bin informs the user how to release the mouse, so that the appropriate causal chains are triggered inside the computer, resulting in the deletion of the file. The icons inform an effective perception-action loop, without themselves having any causal powers in the computer.

To the extent that a user interface succeeds in providing friendly formatting, concealed causality, and clued conduct, it will also offer *ostensible objectivity*. Most of the time the user can simply act as if the interface itself is the total reality of the computer. Indeed some users might be fooled into assuming that the interface is the total reality. We hear humorous stories of a child or grandparent who wondered why an unwieldy box was attached to the screen. Only for more sophisticated purposes, such as debugging a program or repairing hardware, does the limitation of this illusion become inescapable.

Virtual worlds

Suppose you and a friend play virtual tennis at an arcade. You don your helmet and bodysuit, and find yourself in Roland-Garros Stadium, home of the French Open. After admiring the clay court and stadium, you serve to open the first set, and are soon immersed in play.

The stadium, court, net, ball, and racquet that you experience are all, of course, part of a sophisticated user interface, one that exhibits the four qualities described in the last section.

First, it sports friendly formatting. You see red clay, a yellow ball, a graphite tennis racquet, and a green stadium. These are much easier to interpret and use than the complex supercomputer and megabytes of software that control the game.

It conceals causality and clues conduct. When you hit that killer drop volley, it might appear that the head of the racquet caused the ball to sneak across the net. But of course the racquet and ball are just pixels in the user interface, and send no signals back to the supercomputer. The racquet and ball serve only to inform your actions and these, transmitted back via the body suit, trigger a complex but hidden causal sequence within the supercomputer, resulting in the proper updating of registers corresponding to the positions of racquet and ball. A good programmer could update these registers directly. But this would be so slow and cumbersome that even the deftest programmer would lose the match to a modestly talented player who simply acted on the user interface. That is the power, and purpose, of the interface.

Finally, the commercial success of the game depends, in large part, on its ostensible objectivity. Customers want to play tennis, blissfully ignorant of the supercomputer and software hard at work in a back room. Tennis is, for them, the reality. Nothing in their tennis reality resembles the hidden supercomputer, the true causal nexus that makes the game possible. Customers can play as if the tennis ball and racquet had causal powers, even though this is merely a convenient, and entertaining, fiction.

Perception as a multimodal user interface (MUI)

I reject the hypothesis of faithful depiction, the hypothesis that a goal of perception is to match or approximate properties of an objective physical world.

Instead I propose the MUI hypothesis: *The conscious perceptual experiences of an agent are a multimodal user interface between that agent and an objective world*.

To say here that a world is objective means that the world's existence does not depend on that agent. MUI theory makes no ontological claims about the nature of that objective world, nor does it claim any match or resemblance between properties of the interface and properties of the world. They can be as different as tennis balls and integrated circuits. In this regard, the MUI hypothesis is more conservative, i.e., makes fewer claims, than the hypothesis of faithful depiction.

If you have a conscious experience of a rock or a tree, the hypothesis of faithful depiction claims that, if your experience is not illusory, then there must be a rock or tree in the objective world whose properties at least roughly match those of your experience. MUI theory is not committed to this claim, even if your experience is not illusory. It allows countless possibilities for what entities or properties in the objective world might have triggered your perceptual experience of a rock or a tree. Chances are, for a successful perceptual interface, there is no match between properties of conscious experience and properties of the objective world. Instead your perceptual experiences are, in the typical case, substantially less complex and in an entirely different format than the objective properties that trigger them. It is this failure to match, due to adaptive simplification and reformatting, that contributes to the success and usefulness of perceptual experiences. Concern about whether perception is veridical is, within the MUI framework, a category error. The proper concern is whether perception usefully informs action.

According to MUI theory, the objects of our everyday experience—tables, chairs, mountains, and the moon—are not public. If, for instance, I hand you a glass of water, it is natural, but false, to assume that the glass I once held is the same as, i.e., numerically identical with, the glass you now hold. Instead, according to MUI theory, the glass I held was, when I observed it, an icon of my perceptual experience within my MUI, and the glass you now hold, when you observe it, is an icon of your MUI, and they are numerically distinct. There are two glasses of water, not one. And if a third person watches the transaction, there are three glasses of water, not one.

This claim seems, to most, absurd, and subject to straightforward refutation. Searle (2003, p. 275ff), for instance, argues against the denial of public physical objects as follows: First, we all assume, quite naturally, that we at least sometimes communicate successfully with each other. This requires that we have publicly available meanings in a public language, so that we can both mean, or intend, the same thing by utterances such as “this glass of water”. But this requires that we have publicly available objects of reference, e.g., a publicly available glass of water, so that when I say “this glass of water” I am referring to the same object as you do when you say “this glass of water”. But this implies that we both share perceptual access to the same object, which makes it a public object. Thus, concludes Searle, there are public physical objects and the correct philosophy of perception is direct realism.

This argument is easily seen false by counterexample. Bob and Tom, playing virtual tennis, can talk meaningfully about “the tennis ball” they are hitting back and forth; they can agree, for instance, that Tom hit “the tennis ball” out of court, thus losing a point. There is, patently, no public tennis ball. Instead, a supercomputer in the back room feeds signals to the helmet displays of Bob and Tom and each, in consequence, constructs his own tennis-ball experience. But Bob’s tennis-ball experience is numerically distinct from Tom’s. And there is no other tennis ball around to serve the role of public tennis ball. Thus public physical objects are not required for meaningful communication, and Searle’s argument for public physical objects, and direct realism, fails.

This counterexample is instructive, for it shows precisely why Searle’s argument fails. Bob and Tom can speak meaningfully about “the tennis ball” because their experiences are properly coordinated. Searle’s argument assumes that such coordination *requires* a public tennis ball. But this assumption is false: the coordination in the counterexample is accomplished not by a public tennis ball, but by a hidden supercomputer. There are, of course, countless other ways such coordination could occur. The supercomputer is simply one example to help free the imagination from the straight jacket of faithful depiction and direct realism. Those entities in the objective world that allow our tennis ball experiences to be coordinated need not themselves resemble, in any way, a tennis ball.

Again, according to MUI theory, everyday objects such as tables, chairs and the moon exist only as experiences of conscious observers. The moon I experience only exists when I look, and the moon you experience only exists when you look. We never see the same moon. We only see the moon icons we each construct each time we look.

This claim sounds, to most, absurd, and easily refuted. There are several arguments for its absurdity.

First, that chair can’t exist only when I look at it. For I can look away and still touch it. So it still exists. Or I can look away and you can look at it, and confirm to me that it is still there. So again it still exists.

But this argument is easily refuted by the virtual-tennis counterexample. Bob can claim that the tennis ball he and Tom are playing with doesn’t just exist when he looks at it. After all, he can look away and still touch the tennis ball. Or he can look away and Tom can look at it. So, Bob can claim, the tennis ball still exists even when he doesn't look at it. But Bob's claim is patently false.

A second argument: If you think that this train thundering down the tracks is just an icon of your user interface, and doesn't exist when you don't perceive it, then why don't you step in front of it? You'll soon find out that it's more than an icon. And I will see, after you are gone, that it still exists.

This argument makes an elementary confusion. The train, according to MUI theory, is an icon that you are triggered to construct when you interact with some aspect of the object world, an aspect that is not, itself, a train and doesn't resemble a train. So, according to MUI theory, the train icon exists only when you perceive it, and you should not take it literally, i.e., as resembling an element of the objective world. But taking something literally is different than taking it seriously. If your MUI is functioning properly, you should take its icons seriously, but not literally. The point of the icons is that they inform your behavior in ways that are adaptive to your niche. Creatures that don't take their well-adapted icons seriously have a pathetic habit of going extinct. The train icon usefully informs your behaviors, including such laudable behaviors as staying off of train-track icons. Similarly, on your windows interface, you take icons seriously but not literally. Just because a file icon does not literally resemble a file, you don't willy nilly drag the icon to the recycle bin, for you might lose weeks of work. So the MUI theorist is careful about stepping before trains for the same reason that computer users are careful about dragging icons to the recycle bin.

A third argument highlights the stubbornness of icons. Look, if that wall is just an icon I construct, why can't I walk through it? Shouldn't it do what I want if I construct it?

Not at all. You construct the subjective necker cube that you see in Figure 1. But it doesn't do everything you want, whenever you want. For instance, sometimes you see a cube with corner A in front and sometimes a different cube with corner B in front. But try to make yourself switch, at will and instantly, between the two cubes and you will find that your cube constructions are stubborn. They don't always do what you want when you want. Or, if you are good at switching between the two cubes, then try to see the edges of the cube as wiggly rather than straight. No chance. The fact that we construct our icons does not entail that they are subject to our every whim. We are triggered to construct icons by our interactions with the objective world (whatever its nature might be) and, once so triggered, we construct our icons according to certain probabilistic rules (see, e.g, Hoffman, 1998). The objective world and our rules for icon construction make the icons stubborn. Still, these icons exist only in our conscious perceptions.

A fourth argument drops naïve realism in favor of sophisticated realism. We grant, so the argument goes, that everyday objects such as tables, chairs and the moon are just our icons, and exist only in our conscious experiences. But what's new? Physicists have long told us that the apparent solidity of a table is an illusion. The table is mostly empty space with atoms, quarks, leptons, and myriads of other subatomic particles darting about probabilistically. Our perception of a table's surface simply approximates the envelope of this probabilistic activity, and in this sense the hypothesis of faithful depiction is in fact correct. There are no objective tables, just objective particles.

The mistake here is analogous to a computer user who admits that file icons on the display are just conventional symbols, not the actual files, but then puts a magnifying glass over an icon, sees its pixels, and concludes that these pixels are the actual file. File icons are indeed composed of pixels, but these pixels are part of the interface, not elements of the file itself. Similarly, tables are indeed composed of atoms and quarks, but atoms and quarks are part of the MUI, not elements of the objective world. The MUI may be hierarchically organized, but different levels of this hierarchy are part of the MUI, not of the objective world.

Placing atoms and subatomic particles in the MUI rather than in the objective world is compatible with quantum theory. Indeed, the Copenhagen Interpretation of quantum theory asserts that the dynamical properties of such particles have real values only in the act of observation (see, e.g., Albert, 1992; Wheeler & Zurek, 1983, Zurek, 1989). That is, they are part of the observer's MUI. Quantum physics does not contradict MUI theory.

A fifth argument is sociological. Ideas similar to MUI theory have been around in various forms of idealism. But, as Searle (2004, p. 48) says, "Idealism had a prodigious influence in philosophy, literally for centuries, but as far as I can tell it has been as dead as a doornail among nearly all the philosophers whose opinions I respect, for many decades, so I will not say much about it."

One could reply that a similar argument, centuries ago, could have dismissed a spherical earth: everyone respectable thought it flat. But this reply does not go far enough. Idealism was tried and rejected. Why turn back the clock with MUI theory?

This is a simple misunderstanding. MUI theory is not idealism. It does not claim that all that exists are conscious perceptions. It claims that our conscious perceptions need not resemble the objective world, whatever nature the objective world might happen to take. MUI theory is compatible with a physicalist ontology, but MUI theory is not itself committed to any particular ontology. For all we know, we could be the lucky species, the one species in 70 million whose perceptual experiences just happen to resemble the true nature of an objective physical reality. Long odds, but not impossible. MUI theory does not, by itself, rule it out. It simply invites us to take a sober look at the odds.

Conscious realism

MUI theory, we have seen, makes no claim about the nature of the objective world. In this section I propose a theory that does: *conscious realism*. One could accept MUI theory and reject conscious realism, or reject both. But they fit well, and together provide a novel solution to the mind-body problem. Conscious realism is a proposed answer to the top question in the list of 125 questions posed by *Science*: "What is the universe made of?"

Conscious realism asserts the following: *The objective world, i.e., the world whose existence does not depend on the perceptions of a particular conscious agent, consists entirely of conscious agents.*

To make conscious realism precise I give, in the next section, a mathematical definition of *conscious agent*. For now, I describe conscious realism less formally and contrast it with other theories.

First, conscious realism is a nonphysicalist monism. What exists in the objective world, independent of my perceptions, is a world of conscious agents, not a world of unconscious particles and fields. Those particles and fields are icons in the MUIs of conscious agents, but are not themselves fundamental denizens of the objective world. Consciousness is fundamental. It is not a latecomer in the evolutionary history of the universe, arising from complex interactions of unconscious matter and fields. Consciousness is first; matter and fields depend on it for their very existence.

According to conscious realism, when I visually experience a table, I interact with a system, or systems, of conscious agents, and represent that interaction in my conscious experience as a table icon. Admittedly, the table gives me little insight into those conscious agents and their dynamics. The table is a dumbed-down icon, adapted to my needs as a member of a species in a particular niche, but not necessarily adapted to give me insight into the true nature of the objective world that triggers my construction of the table icon. When, however, I see you, I again interact with a conscious agent, or a system of conscious agents. And here my icons give deeper insight into the objective world: they convey that I am, in fact, interacting with a conscious agent, namely you.

Conscious realism is not panpsychism nor entails panpsychism. Panpsychism claims that all objects, from tables and chairs to the sun and moon, are themselves conscious (Hartshorne, 1937/1968; Whitehead, 1929/1979), or that many objects, such as trees and atoms, but perhaps not tables and chairs, are conscious (Griffin, 1998). Conscious realism, together with MUI theory, claims that tables and chairs are icons in the MUIs of conscious agents, and thus that they are conscious experiences of those agents. It does not claim, nor entail, that tables and chairs are themselves conscious or conscious agents. By comparison, to claim, in the virtual-tennis example, that a supercomputer is the objective reality behind a tennis-ball icon is not the same as to claim that the tennis-ball icon is itself a supercomputer. The former claim is, for purposes of the example, true but the latter claim is clearly false.

Conscious realism is not the transcendental idealism of Kant (1781/2003). Exegesis of Kant is notoriously difficult and controversial. The standard interpretation has him claiming, as Strawson (1966, p. 38) puts it, that "reality is supersensible and that we can have no knowledge of it." We cannot know or describe objects as they are in themselves, the noumenal objects, we can only know objects as they appear to us, the phenomenal objects (see also Prichard, 1909). This interpretation of Kant precludes any science of the noumenal, for if we cannot describe the noumenal then we cannot build scientific theories of it. Conscious realism, by contrast, offers a scientific theory of the noumenal, viz., a mathematical formulation of conscious agents and their interactions. This difference between Kant and conscious realism is, for the scientist, fundamental. It is the difference between doing science or not doing science. This fundamental difference also holds for other interpretations of Kant, such as that of Allison (1983).

Many interpretations of Kant have him claiming that the sun and planets, tables and chairs, are not mind independent, but depend for their existence on our perception. With this claim of Kant, conscious realism and MUI theory agree. Of course many current theorists disagree. For instance, Stroud (2000, p. 196), discussing Kant, says, "It is not easy to accept, or even to understand, this philosophical theory. Accepting it presumably means believing that the sun and the planets and the mountains on earth and everything else that has been here so much longer than we have are nonetheless in some way or other dependent on the possibility of human thought and experience. What we thought was an independent world would turn out on this view not to be fully independent after all. It is difficult, to say the least, to understand a way in which that could be true."

But it is straightforward to understand a way in which that could be true. There is indeed something that has been here so much longer than we have. But that something is not the sun and the planets and the mountains on earth. It is dynamical systems of interacting conscious agents. The sun and planets and mountains are simply the icons of our MUI that we are triggered to construct when we interact with these dynamical systems. The sun you see is a momentary icon, constructed on the fly each time you experience it. Your sun icon does not match or approximate the objective reality that triggers you to construct a sun icon. It is a species-specific adaptation, a quick and dirty guide, not an insight into the objective nature of the world.

Mathematical definition of *conscious agent*

The ontology of conscious realism proposed here rests crucially on the notion of *conscious agent*. The reader might reasonably ask if this notion is just intuitive, or if it can be made precise. After all, if we seek a scientific theory of the mind-body problem, then mathematical rigor is required. In this section I present and discuss a mathematical definition of conscious agent. This presentation, due to limitations of space, is necessarily brief. More extensive mathematical treatments are available (Bennett, Hoffman, & Prakash, 1989, 1991; Bennett, Hoffman, & Kakarala, 1993; Bennett, Hoffman, & Murthy, 1993; Bennett et al, 1996). Those readers not interested in the mathematical details can safely skip this section. I begin with the definition of a conscious observer (Bennett et al, 1989, p. 23).

Definition 1 (Conscious observer). A *conscious observer* is a six-tuple, $((X, \mathcal{X}), (Y, \mathcal{Y}), E, S, p, q)$, satisfying the following conditions: (1) $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ are measurable spaces with $E \in \mathcal{X}$ and $S \in \mathcal{Y}$; (2) $p: X \to Y$ is a measurable surjective function with $p(E) = S$; (3) Let $(E, \mathcal{E})$ and $(S, \mathcal{S})$ be the measurable spaces on E and S respectively induced from those of X and Y. Then q is a markovian kernel on $S \times \mathcal{E}$ such that, for each $s \in S$, $q(s,\bullet)$ is a probability measure supported in $p^{-1}\{s\} \cap E$.

(Recall that a measurable space is a set, X, and a collection, $\mathcal{X}$, of subsets of X that contains X and is closed under complement and countable union; these subsets are called events and $\mathcal{X}$ is called a σ-algebra. A markovian kernel is an indexed collection of probability measures, satisfying certain technical conditions (see, e.g., Bennett et al, 1989).)

The definition of conscious observer can be understood with a concrete example: Ullman's (1979) theory of seeing 3D object structure from image motion. Human vision has the remarkable ability to construct 3D objects when it views certain displays of 2D motion. This is one reason we see 3D objects when watching television or movies. Ullman's theory is captured by the following theorem:

Ullman's Theorem: Three distinct orthographic views of four noncoplanar points almost surely have no rigid 3D interpretations. If the views do have a rigid 3D interpretation, then almost surely they have two, which are orthographic reflections of each other.

The phrase "almost surely" is a technical term meaning "except possibly for sets having Lebesgue measure zero." A 3D structure is rigid if all distances between all points in the structure do not change over time.

Ullman's theorem guides the construction of 3D objects as follows. If an observer is given three images, and each contains at least four feature points, then the observer can do a simple computation to determine if, in principle, a rigid 3D object could exist which, when viewed from the right directions, would project to the three given images. If the feature points were placed at random, i.e., according to a uniform distribution, on the three images, then this simple computation will, almost surely, yield no solution, thus indicating that no rigid 3D object should be constructed. If, however, the computation does find a solution then, almost surely, it will find two that are mirror reversals of each other, and it will give the coordinates of their 3D structures. The observer thus constructs, and consciously perceives, these 3D structures, one at a time, in alternation. In psychophysical experiments with motion displays of this type, human observers report seeing one of the 3D interpretations for a while, then suddenly switching to seeing the other, mirror-reversed, 3D interpretation.

The set of possible image data for a conscious observer based on Ullman's theorem consists of all possible three images with four feature points each. Since each feature point requires two coordinates to specify its position and there are four points in one image, it requires eight coordinates to specify all the feature points in one image. For three images, then, it requires twenty-four coordinates. Thus the set of all possible three images with four points is a twenty-four dimensional Euclidean space, i.e., $\boldsymbol{R}^{24}$. This $\boldsymbol{R}^{24}$ is the set Y in the definition of conscious observer. It is called the *premise space* of the conscious observer, since it is the space of possible inputs, or premises, for this conscious observer.

Most points of Y have no possible rigid interpretations. However a subset of Y, consisting of those points that do have rigid interpretations, has Lebesgue measure zero in Y. This subset is the set S in the definition of conscious observer, and is called the *special* or *distinguished premises*. Only when given a point of S, a special premise, does this conscious observer construct, and consciously experience, a rigid 3D structure.

A point on a 3D structure requires three coordinates to specify its position. A 3D structure with four points thus needs twelve coordinates to be specified. To specify such a structure at three distinct times thus requires thirty-six coordinates. The set of all possible structures, rigid or not, is thus $\boldsymbol{R}^{36}$. This $\boldsymbol{R}^{36}$ is the set X in the definition of conscious observer. It is called the *conclusion* or *interpretation space* for this conscious observer.

Most structures in X are not rigid. The subset of X consisting of rigid structures is the set E in the definition of conscious observer. It is called the set of *special* or *distinguished interpretations*. The set X provides the syntactic framework in which the set E can be properly described.

The sets X and Y are related by orthographic projection, given by $(x,y,z) \rightarrow (x,y)$. That is, orthographic projection simply strips off the depth coordinate. This induces a map $p: X \rightarrow Y$, which is the map p in the definition of conscious observer, and called the *perspective map*. Note that $p(E)=S$.

Almost no image data $y \in Y$ has rigid 3D interpretations. But any image data $s \in S$ has a rigid 3D interpretation and, according to Ullman's Theorem, generically it has two such interpretations. For each $s \in S$, the markovian kernel q in the definition of conscious observer, called the *interpretation kernel*, gives a probability measure, $q(s,\bullet)$, that is supported on the two rigid interpretations in $p^{-1}\{s\} \cap E$. When a conscious observer is given a premise s, the probability measure $q(s,\bullet)$ describes the *conscious experience* of that observer: The points e_γ in the support of $q(s,\bullet)$ are all the potential conscious experiences of the observer, given the premise s. If there is only one point e in this support, then $q(s,e)=1$ and e is the conscious experience of the observer. If there are two or more points in the support of $q(s,\bullet)$, then the conscious experience of the observer is multistable, and the probability that the conscious experience of the observer is a particular interpretation e_γ is $q(s, e_\gamma)$.

One might object that the set E could simply represent the unconscious states of, say, a robot, and that therefore the definition of conscious observer, despite its name, has nothing to do with consciousness. This objection is an elementary mistake. Using the integers to count apples doesn't preclude using the integers to count oranges.

The definition of conscious observer generalizes the standard Bayesian formulation of perception (Knill & Richards, 1996). According to this formulation, as applied to vision, an observer is given a sequence of images I and wants to compute the probability of various world interpretations W. That is, the observer wants to compute the conditional probability $P(W \mid I)$. By Bayes' theorem, we can write $P(W \mid I) = P(I \mid W)P(W)/P(I)$. The term $P(W \mid I)$ is called the posterior probability, the term $P(W)$ the prior probability, and the term $P(I \mid W)$ the likelihood function. To be well defined, this formulation requires, of course, that $P(I)$ is not zero. The interpretation kernel in Definition 1 of conscious observer relaxes this requirement, and allows one to have posterior probabilities conditioned on sets of measure zero (Bennett et al, 1996).

The definition of conscious observer is here presented, for simplicity, in the noise-free case. It has been generalized to handle noise (Bennett, Hoffman, & Kakarala, 1993).

A conscious observer, as just defined, has conscious experiences but no dynamics, so it does not act in the sense of completing a perception-action loop (see, e.g., Hurley, 1998). For that we turn to the definition of a conscious agent. Intuitively, a conscious agent is a markovian dynamics on a state space whose points are conscious observers. By moving about on this state space, the conscious agent updates how it consciously perceives as a function of how, and what, it currently perceives. This is captured in the following definition.

Definition 2 (Conscious Agent). Let $O = \{O_\alpha \mid \alpha \in \mathcal{I}\}$ be a collection of conscious observers, where $\mathcal{I}$ is any index set. Let $\mathcal{O}$ be a σ-algebra on O. Let S_α denote the special premises for observer O_α, i.e., $S_\alpha = \{s_{\alpha\beta} \mid \beta \in \mathcal{J}_\alpha\}$, where $\mathcal{J}_\alpha$ is an index set for S_α. Let $P = \{S_\alpha \mid \alpha \in \mathcal{I}\}$. A *conscious agent* is a pair, $A = (\mu_0, K)$, where μ_0 is a probability measure on $(O, \mathcal{O})$, and K is a markovian kernel on $P \times \mathcal{O}$.

The kernel K is called the *action kernel* of the conscious agent. It describes probabilistically how the conscious agent acts on its conscious experiences. The dynamics of conscious agents has been studied in some detail, and used to derive quantum theory (Bennett et al, 1989, chapter 10).

With these definitions we can now state precisely the relation between conscious agents and their MUIs: A conscious agent is an entire markovian dynamics on a state space of conscious observers; the MUI of that conscious agent is the set of distinguished interpretations, E_α, for the conscious observers O_α in its state space.

The mistake made by physicalist approaches to the mind-body problem, and to scientific problems more generally, is to assume that the sets E_α describe structures of the objective, i.e., mind-independent, world. Physicalist approaches to the mind-body problem try to bootstrap conscious agents from the E_α alone. But this is destined to fail. One cannot get the structure of a conscious agent, as given in Definition 2, by trying to bootstrap up from the sets E_α. They will not, by themselves, give you the other spaces, maps and kernels that constitute a conscious agent. Moreover, these sets are only one part of a conscious agent and, a fortiori, they are not independent of that agent. One also easily sees the error of panpsychism. It assumes that points of the sets E_α are themselves conscious, when it is the entire conscious agent, not its sets E_α, that is conscious.

A few implications of the definition of conscious agent should be made explicit. First, a conscious agent is not necessarily a person. All persons are conscious agents, or heterarchies of conscious agents, but not all conscious agents are persons. Second, the experiences of a given conscious agent might be utterly alien to us; they may constitute a modality of experience no human has imagined, much less experienced. Third, the dynamics of conscious agents does not, in general, take place in ordinary four-dimensional space-time. It takes place in state spaces of conscious observers, and for these state spaces the notion of dimension might not even be well defined. Certain conscious agents might employ a four-dimensional space-time as part of their MUI, i.e., as part of the structure of their set E. But again, this is not necessary. From these comments it should be clear that the definition of conscious agent is quite broad in scope. Indeed, it plays the same role for the field of consciousness that the notion of Turing machine plays for the field of computation (Bennett et al, 1989).

The asymptotic behavior of dynamical systems of conscious agents can create new conscious agents (Bennett et al, 1989). But computer simulations of these dynamics will no more create conscious experiences than computer simulations of weather will produce rain.

One reader of this section thought it pointless. Suppose, he said, we accept all the mathematics. Then we have a mathematical definition of the technical term “conscious agent.” But, he said, one must still argue that “conscious agents” are indeed conscious and agents. And that, he said, I have not done.

This is an elementary mistake. Once one has given a mathematical definition of the technical term “stochastic process” one does not then need to argue that “stochastic processes” are indeed stochastic and processes. The mathematical description is just that, a description, and not itself the thing described. The mathematical description, for purposes of science, stands or falls on its ability to generate theories that are insightful and empirically adequate.

The mind-body problem

Having a precise definition of conscious agent, we can now use MUI theory and conscious realism to sketch a solution to the mind-body problem. Exactly what that problem is depends, of course, on one’s assumptions. If one adopts physicalism, then the central scientific problem is the following:

Physicalist Mind-Body Problem: Describe precisely how conscious experience arises from, or is identical to, certain types of physical systems.

As we discussed before, there are, so far, no scientific theories of the physicalist mind-body problem. If, instead, one adopts conscious realism then the central mind-body problem is as follows:

Conscious-Realist Mind-Body Problem: Describe precisely how conscious agents construct physical objects and their properties.

Here there is good news. We have substantial progress on this mind-body problem, and there are real scientific theories. We now have mathematically precise theories about how one type of conscious agent, namely human observers, might construct the visual shapes, colors, textures, and motions of objects (see, e.g., Hoffman, 1998; Knill & Richards, 1996, Palmer, 2000).

For instance, one example we have discussed already is Ullman's (1979) theory of the construction of 3D objects from image motion. This theory is mathematically precise and allows one to build computer vision systems that simulate the construction of such 3D objects. There are many other mathematically precise theories and algorithms for how human observers could, in principle, construct 3D objects from various types of image motions (e.g., Faugeras & Maybank, 1990; Hoffman & Bennett, 1986; Hoffman & Flinchbaugh, 1982; Huang & Lee, 1989; Koenderink & van Doorn, 1991; Longuet-Higgins & Prazdny, 1980). We also have precise theories for constructing 3D objects from stereo (Geiger, Ladendorf & Yuille, 1995; Grimson, 1981; Marr & Poggio, 1979), shading (Horn & Brooks, 1989), and texture (Aloimonos & Swain, 1988; Witkin, 1981). Researchers debate, as they should, the relevance and empirical adequacy of each such theory as a model of human perception. This is just normal science.

Now, admittedly, almost without exception the authors of these theories accept the hypothesis of faithful depiction and conceive of their theories as specifying methods by which human observers can *reconstruct* or approximate the true objective properties of independently existing physical objects. But each of these theories can equally well be reinterpreted simply as specifying a method of object *construction*, not reconstruction. The mathematics is indifferent between the two interpretations. It does not require the hypothesis of independently existing physical objects. It is perfectly compatible with the hypothesis of conscious realism, and the mind dependence of all objects. So interpreted, the large and growing literature in computational vision, and computational perception more generally, is concrete scientific progress on the mind-body problem, as this problem is posed by conscious realism. It gives mathematically precise theories about how certain conscious agents construct their physical worlds. The relationship between the conscious and the physical is thus not a mystery, but the subject of systematic scientific investigation and genuine scientific theories.

What one gives up, to have this scientific progress on the mind-body problem, is the dearly held belief that physical objects and their properties exist independently of the conscious agents that perceive them. Piaget claimed that children, at about nine months of age, acquire object permanence, the belief that physical objects exist even when they are not observed (Piaget, 1954; but see Baillargeon, 1987). Conscious realism claims that object permanence is an illusion. It is a useful and convenient fiction that substitutes for a situation which, for the child, is too subtle to grasp: Something continues to exist when the child stops observing, but that something is not the physical object that the child sees when it observes; that something is, instead, a complex dynamical system of conscious agents that triggers the child to create a physical-object icon when the child interacts with that system. For the child it is much simpler, and rarely problematic, to simply assume that the physical object it perceives is what continues to exist when it does not observe. Indeed, only when one faces the subtleties of, e.g., quantum theory or the mind-body problem, does the utility of the illusion of object permanence finally break down, and a more sophisticated, and comprehensive, ontology become necessary.

With physicalist approaches to the mind-body problem, one faces a difficult question of causality: If conscious experience arises somehow from brain activity, and if the physical world is causally closed, then how, precisely, does conscious experience cause anything? It seems, for instance, that I eat pistachio ice cream because I feel hungry and I like the taste of pistachio. Do my conscious experiences in fact cause my eating behaviors? No, say nonreductive functionalists, such as Chalmers (1995), who claim that functional properties of the brain give rise to, but are not identical with, conscious experiences. Instead they often endorse epiphenomenalism: Brain activity gives rise to conscious experiences but, since the physical realm is causally closed, conscious experiences themselves have no causal consequences. It seems like I eat pistachio because it tastes good, but this is an illusion. Moreover, I believe that I consciously experience the taste of pistachio, but I would believe this whether or not I in fact consciously experience this taste. This is a desperate claim and, as I mentioned before, close to an outright reductio of the position. Reductive functionalists, by contrast, do not endorse epiphenomenalism, since they claim that conscious experiences are *identical* to certain functional states of the brain, and conscious experiences therefore possess the causal properties of those functional states. However, reductive functionalism has recently been disproved by the Scrambling Theorem which shows that, if one assumes only that conscious experiences can be represented mathematically, then conscious experiences and functional relations are not numerically identical (Hoffman, 2006).

Conscious realism leads to a different view of causality, a view I call *epiphysicalism*: Conscious agents are the only locus of causality, and such agents construct physical objects as elements of their MUIs; but physical objects have no causal interactions among themselves, nor any other causal powers. Physical objects, as icons of a conscious agent's MUI, can *inform*, but do not cause, the choices and actions of a conscious agent. When a cue ball hits an eight ball and sends it careening to the corner pocket, the cue ball does not cause the movement of the eight ball any more than the movement of a file icon to the recycle bin causes the bin to open or a file to be deleted. A useful user interface offers, as we have discussed, concealed causality and ostensible objectivity. It allows one to act, in all but the most sophisticated situations, as if the icons had causal powers, and in complete ignorance of the true causal chains. No law of physics describes a causal interaction, because all such laws pertain to the behavior of the contents of interfaces, not to the behavior of conscious agents. The causal behaviors of conscious agents are described by interpretation kernels and action kernels (defined in section 7). The perceptual conclusions of one conscious observer might be among the premises of a second conscious observer and, thereby, inform but not cause the perceptions of the second (Bennett et al, 1989). Attractors in the asymptotic stochastic behavior of a system of conscious agents might be among the premises of other conscious agents and thereby inform, but not cause, their behavior (Bennett et al, 1989).

So, in particular, epiphysicalism entails that the brain has no causal powers. The brain does not cause conscious experience; instead, certain conscious agents, when so triggered by interactions with certain other systems of conscious agents, construct brains (and the rest of human anatomy) as complex icons of their MUIs. The neural correlates of consciousness are many and systematic not because brains cause consciousness, but because brains are useful icons in the MUIs of certain conscious agents. According to conscious realism, you are not just one conscious agent, but a complex heterarchy of interacting conscious agents, which can be called your *instantiation* (Bennett et al, 1989 give a mathematical treatment). One symbol, created when certain conscious agents within this instantiation observe the instantiation, is a brain.

Does this view entail that we should stop the scientific study of neural correlates of consciousness? No. If we wish to understand the complex heterarchy of conscious agents in human instantiations, we must use the data that our MUIs provide, and that data takes the form of brain icons. Brains do not create consciousness; consciousness creates brains as dramatically simplified icons to a realm far more complex, a realm of interacting conscious agents. When, for instance, we stimulate primary visual cortex and see phosphenes, the cortex does not cause the phosphenes. Instead, certain interactions between conscious agents cause the phosphenes, and these interactions we represent, in greatly simplified icons, as electrodes stimulating brains.

Evolution

One objection to conscious realism invokes evolution. We now know, the argument goes, that the universe existed for billions of years before the first forms of life, and probably many millions more before the first flickers of consciousness. Natural selection, and other evolutionary processes first described by Darwin, have since shaped life and consciousness into "endless forms, most beautiful and most wonderful." This contradicts the claim of conscious realism, viz., that consciousness is fundamental and that matter is simply a property of certain icons of conscious agents.

Four responses. First, although it is true that evolutionary theory has been interpreted, almost exclusively, within the framework of a physicalist ontology, the mathematical models of evolution do not require this ontology. They can be applied equally well to systems of conscious agents and, indeed, such an application of evolutionary game theory (Maynard-Smith, 1982; Skyrms, 2000) is quite natural. Systems of conscious agents can undergo stochastic evolution, and conscious agents can be synthesized or destroyed in the process (Bennett et al, 1989, 2002). There is simply no principled reason why evolution requires physicalism. Evolutionary changes in genes and body morphology can be modeled by evolution whether those genes and bodies are viewed as mind independent or mind dependent. The mathematics doesn't care.

Nor does the fossil evidence. A dinosaur bone dated to the Jurassic can be interpreted along physicalist lines as a mind-independent object or, with equal ease, as a mind-dependent icon that we construct whenever we interact with a certain long-existing system of conscious agents. For the conscious realist there is, no doubt, interesting and fundamental work to be done here: We want a rigorous mathematical theory of the evolution of conscious agents that has the property that, when this evolution is projected onto the relevant MUIs, it gives us back the current physicalist model of evolution. That is, we must exhibit physicalist evolutionary models as special cases, in fact projections, of a richer and more comprehensive evolutionary theory. But this is nothing special about evolution. We want the same for all branches of science. For instance we want, where possible, to exhibit current laws of physics as projections of more general laws or dynamics of conscious agents. Some current laws of physics, or of other sciences, might be superceded or discarded as the science of conscious realism advances, but those that survive should be exhibited as limiting cases or projections of the more complete laws governing conscious agents and their MUIs.

Second, according to conscious realism it simply is not true that consciousness is a latecomer in the history of the universe. Consciousness has always been fundamental, and matter derivative. The picture of an evolving unconscious universe of space-time, matter and fields that, over billions of years, fitfully gives birth first to life, then to consciousness, is false. The great psychological plausibility of this false picture derives from our penchant to commit a reification fallacy, to assume that the icons we create are in fact objects independent of us and fundamental in the universe. We embrace this fallacy because our MUI successfully informs our behavior and has ostensible objectivity, because we construct the icons of our MUI so quickly and efficiently that most of us never discover that we in fact construct them, and because we first commit the fallacy in infancy and are rarely, if ever, encouraged to challenge it. The illusion of object permanence starts by nine months, and does not go easy.

Third, standard evolutionary theory itself undercuts the reification fallacy that underlies the hypothesis of faithful depiction. Natural selection prunes perceptual systems that do not usefully guide behavior for survival; natural selection does *not* prune perceptual systems because they don't match or approximate objective reality (see, e.g., Radnitzky & Bartley, 1987). The perceptual systems of roaches, we suspect, give little insight into the complexities of objective reality. The same for lice, maggots, nematodes and an endless list of creatures that thrived long before the first hominoid appeared and will probably endure long after the last expires. Perceptual systems arise without justification from random mutations and, for 99 percent of all species that have sojourned the earth, without justification they have disappeared in extinction. The perceptual icons of a creature must quickly and successfully guide its behavior in its niche, but they need not give truth. The race is to the swift, not to the correct. As Pinker (1997, p. 561) puts it, "We are organisms, not angels, and our minds are organs, not pipelines to the truth. Our minds evolved by natural selection to solve problems that were life-and-death matters to our ancestors, not to commune with correctness…"

Shepard hopes otherwise: "Possibly we can aspire to a science of mind that, by virtue of the evolutionary internalization of universal regularities in the world, partakes of some of the mathematical elegance and generality of theories of that world." (2001, p. 601). It is, one must admit, *logically* possible that the perceptual icons of *homo sapiens*, shaped by natural selection to permit survival in a niche, might also just happen to faithfully represent some true objects and properties of the objective world. But this would be a probabilistic miracle, a cosmic jackpot against odds dwarfing those of the state lottery. The smart money is on humble icons with no pretense to objectivity.

But this last response might not go far enough, for it grants that natural selection, understood within a physicalist framework, can shape conscious experience. Perhaps it cannot. Natural selection prunes functional propensities of an organism relevant to its reproductive success. But the Scrambling Theorem proves that conscious experiences are not identical with functional propensities (Hoffman, 2006). Thus natural selection acting on functional propensities does not, ipso facto, act as well on conscious experiences. A nonreductive functionalist might counter that, although conscious experiences are not identical to functional properties, nevertheless conscious experiences are caused by functional properties, and thus are subject to shaping by natural selection. The problem with this, as we have discussed, is that no one has come close to turning the idea of nonreductive functionalism into a genuine scientific theory, and the failure appears to be principled. Moreover the idea itself, together with the assumption of the causal closure of the physical, entails epiphenomenalism with its implication that my beliefs about my conscious experiences are not caused by my experiences and would not change even if I had no experiences. Arguably a reductio of the position. Thus the burden of proof is clearly on those who wish to claim that natural selection, understood within a physicalist framework, can shape conscious experience. Understood within the framework of conscious realism, natural selection has no such obstructions to shaping conscious experiences.

Conclusion

Abraham Pais, describing his interactions with Einstein, wrote "Einstein never ceased to ponder the meaning of the quantum theory … We often discussed his notions on objective reality. I recall that during one walk Einstein suddenly stopped, turned to me and asked whether I really believed that the moon exists only when I look at it." (Pais, 1979, p. 907).

MUI theory says that the moon you see is, like any physical object you see, an icon constructed by your visual system. Perception is not objective reporting but active construction. A perceptual construction lasts only so long as you look, and then is replaced by new constructions as you look elsewhere. Thus the answer to Einstein's question, according to MUI theory, is that the moon you see only exists when you look at it. Of course the moon Jack sees might continue to exist even when the moon Jill sees ceases to exist because she closes her eyes. But the moon Jack sees is not numerically identical to the moon Jill sees. Jack sees his moon, Jill sees hers. There is no public moon.

Something does exist whether or not you look at the moon, and that something triggers your visual system to construct a moon icon. But that something that exists independent of you is not the moon. The moon is an icon of your MUI, and therefore depends on your perception for its existence. The something that exists independent of your perceptions is always, according to conscious realism, systems of conscious agents. Consciousness is fundamental in the universe, not a fitfully emerging latecomer contorting the senseless face of matter.

The mind-body problem is, for the physicalist, the problem of getting consciousness to arise from biology. So far no one has come remotely close to building a scientific theory of how this might happen. This failure is so striking that it leads some to wonder if *homo sapiens* lacks the necessary conceptual apparatus.

For the conscious realist, the mind-body problem is how, precisely, conscious agents create physical objects and properties. Here we have a vast and mathematically precise scientific literature, with successful implementations in computer vision systems.

To a physicalist, the conscious-realist mind-body problem might appear to be a bait and switch that dodges hard and interesting questions: What is consciousness for? When and how did it arise in evolution? How does it now arise from brain activity? Now, admittedly, with conscious realism there is a switch, from the ontology of physicalism to the ontology of conscious realism. This switch changes the relevant questions. Consciousness is fundamental. So to ask what consciousness is for is to ask why something exists rather than nothing. To ask how consciousness arose in a physicalist evolution is mistaken. Instead we ask how the dynamics of conscious agents, when projected onto appropriate MUIs, yields current evolutionary theory as a special case. To ask how consciousness arises from brain activity is also mistaken. Brains are complex icons representing heterarchies of interacting conscious agents. So instead we ask how neurobiology serves as a user interface to such heterarchies. Conscious realism, it is true, dodges some tough mysteries posed by physicalism, but it replaces them with new, and equally engaging, scientific problems.

Nobody explains everything. If you want to solve the mind-body problem you can take the physical as given and explain the genesis of conscious experience, or take conscious experience as given and explain the genesis of the physical. Explaining the genesis of conscious experience from the physical has proved, so far, intractable. Explaining the genesis of the physical from conscious experience has proved quite feasible. This is good news: We don't need a new conceptual apparatus to transform the mind-body problem from a mystery to a routine scientific subject, we just need a change in the direction in which we seek an explanation. We can start with a mathematically precise theory of conscious agents and their interactions. We can, according to the norms of methodological naturalism, devise and test theories of how conscious agents construct physical objects and their properties, even space and time themselves. In the process we need relinquish no method or result of physicalist science, but instead we aim to exhibit each such result as a special case in a more comprehensive, conscious realist, framework.

References

Alais, D., & Blake, R. (Eds.). (2004). *Binocular rivalry.* Cambridge, MA: MIT Press.

Albert, D. (1992). *Quantum mechanics and experience.* Cambridge, MA: Harvard University Press.

Allison, H. E. (1983). *Kant's transcendental idealism: An interpretation and defense.* New Haven: Yale University Press.

Aloimonos, Y., & Swain, M. J. (1988). Shape from texture. *Biological Cybernetics, 58,* 345–360.

Baars, B. (1988). *A cognitive theory of consciousness.* Cambridge, UK: Cambridge University Press.

Baillargeon, R. (1987). Object permanence in 3½- and 4½-month-old infants. *Developmental Psychology, 23,* 655–664.

Bennett, B.M., Hoffman, D.D., Kakarala, R. (1993). Modeling performance in observer theory. *Journal of Mathematical Psychology*, 37, 2, 220–240.

Bennett, B.M., Hoffman, D.D., Murthy, P. (1993). Lebesgue logic for probabilistic reasoning, and some applications to perception. *Journal of Mathematical Psychology*, 37, 1, 63–103.

Bennett, B.M., Hoffman, D.D., Prakash, C. (1989). *Observer mechanics: A formal theory of perception*. San Diego: Academic Press.
— available online free: http://www.cogsci.uci.edu/%7Eddhoff/ompref.html

Bennett, B.M., Hoffman, D.D., Prakash, C. (1991). Unity of perception. *Cognition*, 38, 295–334.

Bennett, B.M., Hoffman, D.D., Prakash, C. (2002). Perception and evolution. In D. Heyer and R. Mausfeld (Eds) *Perception and the physical world: Psychological and philosophical issues in perception.* West Sussex, UK: Wiley, pp. 229–245.

Bennett, B.M., Hoffman, D.D., Prakash, C., Richman, S. 1996. Observer theory, Bayes theory, and psychophysics. In D. Knill and W. Richards (Eds) *Perception as Bayesian inference*. Cambridge University Press, pp. 163–212.

Broad, C. D. (1925). *The mind and its place in nature.* London: Routledge & Kegan Paul.

Brown, R. J., & Norcia, A. M. (1997). A method for investigating binocular rivalry in real-time with the steady-state VEP. *Vision Research, 37,* 2401–2408.

Celesia, G. G., Bushnell, D., Cone-Toleikis, S., & Brigell, M.G. (1991). Cortical blindness and residual vision: Is the second visual system in humans capable of more than rudimentary visual perception? *Neurology, 41,* 862–869.

Chalmers, D. J. (1995). *The conscious mind: In search of a fundamental theory.* Oxford: Oxford University Press.

Chomsky, N. (1980). *Rules and representations.* New York: Columbia University Press.

Chomsky, N. (2000). *New horizons in the study of language and mind.* Cambridge, UK: Cambridge University Press.

Churchland, P.M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*, 67–90.

Churchland, P.S. (1986). *Neurophilosophy: Toward a unified science of the mind/brain.* Cambridge, MA: MIT Press.

Collins, M. (1925). *Colour-blindness.* New York: Harcourt, Brace & Co.

Crick, F. (1994). *The astonishing hypothesis: The scientific search for the soul.* New York: Scribners.

Crick, F., & Koch, C. (1990). Toward a neurobiological theory of consciousness. *Seminars in the Neurosciences, 2,* 263–275.

Critchley, M. (1965). Acquired anomalies of colour perception of central origin. *Brain, 88,* 711–724.

Davidson, D. (1970). Mental events. In L. Foster & J. Swanson (Eds.), *Experience and theory* (pp. 79–101). New York: Humanities Press.

Dennett, D. (1978). Why you can't make a computer that feels pain. In *Brainstorms* (pp. 190–229). Cambridge, MA: MIT Press.

Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection.* New York: Basic Books.

Edelman, G. M. (2004). *Wider than the sky: The phenomenal gift of consciousness.* New Haven, CT: Yale University Press.

Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination.* New York: Basic Books.

Faugeras, O. D., & Maybank, S. (1990). Motion from point matches: Multiplicity of solutions. *International Journal of Computer Vision, 4,* 225–246.

Fodor, J.A. (1974). Special sciences: Or, the disunity of science as a working hypothesis. *Synthese, 28,* 97–115.

Fodor, J. A., & Pylyshyn, Z. (1981). How direct is visual perception? Some reflections on Gibson's 'Ecological Approach'. *Cognition, 9,* 139–196.

Geiger, D., Ladendorf, B., & Yuille, A. (1995). Occlusions and binocular stereo. *International Journal of Computer Vision, 14,* 211–226.

Gibson, J. J. (1950). *The perception of the visual world*. New York: Houghton-Mifflin.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. New York: Houghton-Mifflin.

Gibson, J.J. (1979/1986). *The ecological approach to visual perception.* Hillsdale, NJ: Lawrence Erlbaum Publishers.

Griffin, D.R, 1998. *Unsnarling the world knot: Consciousness, freedom, and the mind-body problem*. Berkeley, CA: University of California Press.

Grimson, W. E. L. (1981). *From images to surfaces.* Cambridge, MA: MIT Press.

Hameroff, S. R., & Penrose, R. (1996). Conscious events as orchestrated space-time selections. *Journal of Consciousness Studies, 3,* 36–53.

Hartshorne, C. (1937/1968). *Beyond humanism: Essays in the philosophy of nature.* Lincoln: University of Nebraska Press.

Helmholtz, H.L.F. v. (1910/1962). *Handbook of physiological optics, Volume III.* New York: Dover.

Hoffman, D. D. (1983). The interpretation of visual illusions. *Scientific American, 249,* 154–162.

Hoffman, D. D. (1998). *Visual intelligence: How we create what we see.* New York: W. W. Norton.

Hoffman, D. D. (2003). Does perception replicate the external world? *Behavioral and Brain Sciences, 26,* 415–416.

Hoffman, D. D. (2006). The scrambling theorem: A simple proof of the logical possibility of spectrum inversion. *Consciousness & Cognition (in press).*

Hoffman, D. D., & Bennett, B. M. (1986). The computation of structure from fixed-axis motion: rigid structures. *Biological Cybernetics, 54*, 71–83.

Hoffman, D. D., & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics, 42,* 197–204.

Horn, B. K. P., & Brooks, M. (Eds.) (1989). *Shape from shading.* Cambridge, MA: MIT Press.

Horton, J. C., & Hoyt, W. F. (1991). Quadratic visual field defects: A hallmark of lesions in extrastriate (V2/V3) cortex. *Brain, 114,* 1703–1718.

Huang, T., & Lee, C. (1989). Motion and structure from orthographic projections. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11,* 536–540.

Hurley, S. L. (1998). *Consciousness in action.* Harvard, MA: Harvard University Press.

Huxley, T. J. (1866). Lessons in elementary psychology, 8, 210.

Kant, I. 1781/2003. *Critique of pure reason*. New York: Dover.

Kim, J. (1993). *Supervenience and mind.* Cambridge, UK: Cambridge University Press.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach.* Englewood, CO: Roberts & Company.

Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A, 8,* 377–385.

Lehar, S. (2003). Gestalt isomorphism and the primacy of subjective conscious experience: A Gestalt Bubble model. *Behavioral and Brain Sciences, 26,* 375–444.

Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature, 379,* 549–553.

Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London, Series B, 208,* 385–397.

Lumer, E. D., Friston, K. J., & Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science, 280,* 1930–1934.

Lycan, W. G. (2003). Chomsky on the mind-body problem. In L. Anthony & N. Hornstein (Eds.), *Chomsky and his critics* (pp. 11–28). Oxford: Blackwell Publishers.

Marr, D. (1982). *Vision*. San Francisco: Freeman Press.

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London B, 204,* 301–328.

McGinn, C. (1989). Can we solve the mind-body problem? *Mind, 98,* 349–366.

Miller, G. (2005). What is the biological basis of consciousness? *Science, 309,* 79.

Noë, A, & Regan, J.K. (2002). On the brain-basis of visual consciousness: A sensorimotor account. In A. Noë & E. Thompson (Eds.), *Vision and mind: Selected readings in the philosophy of perception* (pp. 567-598). Cambridge, MA: MIT Press.

Pais, A. (1979). Einstein and the quantum theory. *Reviews of Modern Physics, 51,* 863–914.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology.* Cambridge, MA: MIT Press.

Place, U. T. (1956). Is consciousness a brain process? *British Journal of Psychology, 45,* 243–255.

Penrose, R. (1994). *Shadows of the mind.* Oxford, UK: Oxford University Press.

Piaget, J. (1954). *The construction of reality in the child.* New York: Basic.

Pinker, S. (1997). *How the mind works.* New York: W. W. Norton & Co.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, *317*, 314–319.

Prichard, H. A. (1909). *Kant's theory of knowledge.* Oxford: Clarendon Press.

Purves, D., & Lotto, R. B. (2003). *Why we see what we do: An empirical theory of vision*. Sunderland, MA: Sinauer.

Radnitsky, G., & Bartley, W. W. (Eds) (1987). *Evolutionary epistemology, rationality, and the sociology of knowledge.* La Salle, Illinois: Open Court.

Rees, G., Kreiman, G., & Koch, C. (2002). Neural correlates of consciousness in humans. *Nature Reviews Neuroscience, 3,* 261–270.

Rizzo, M., Nawrot, M., & Zihl, J. (1995). Motion and shape perception in cerebral akinetopsia. *Brain, 118,* 1105–1127.

Sacks, O. (1995). *An anthropologist on Mars.* New York: Vintage Books.

Searle, J. R. (1992). *The rediscovery of the mind.* Cambridge, MA: MIT Press.

Searle, J. R. (2004). *Mind: A brief introduction.* Oxford, UK: Oxford University Press.

Schneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction*. Reading, MA: Addison-Wesley.

Shepard, R. (2001). Perceptual-cognitive universals as reflections of the world. *Behavioral and Brain Sciences, 24,* 581–601.

Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review, 68,* 141–156.

Smart, J. J. C. (1978). The content of physicalism. *Philosophical Quarterly, 28,* 339–341.

Stapp, H. P. (1993). *Mind, matter, and quantum mechanics.* Heidelberg: Springer-Verlag.

Stapp, H. P. (1996). The hard problem: A quantum approach. *Journal of Consciousness Studies, 3,* 194–210.

Stoffregen, T.A., & Bardy, B. G. (2001). On specification and the senses. *Behavioral and Brain Sciences, 24,* 195–261.

Strawson, P. F. (1966). *The bounds of sense, an essay on Kant's Critique of Pure Reason.* London: Methuen.

Stroud, B. (2000). *The quest for reality: Subjectivism and the metaphysics of colour.* Oxford: Oxford University Press.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21,* 753–759.

Tononi, G., & Sporns, O. (2003). Measuring information integration. *BMC Neuroscience, 4,* 31–50.

Tononi, G., Srinivasan, R., Russell, D. P., & Edelman, G. M. (1998). Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *Proceedings of the National Academy of Sciences of the United States of America, 95, 3198*–3203.

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, MA: MIT Press.

Ullman, S. (1980). Against direct perception. *Behavioral and Brain Sciences, 3,* 373–415.

Wheeler, J. A. & Zurek, W. H. (1983). *Quantum theory and measurement.* Princeton, NJ: Princeton University Press.

Whitehead, A.N. (1929/1979). *Process and reality: An essay in cosmology*. New York: Free Press.

Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence, 17,* 17–45.

Yuille, A., & Buelthoff, H. (1996). Bayesian decision theory and psychophysics. In D. Knill & W. Richards (Eds.) *Perception as Bayesian inference* (pp. 123-161), Cambridge , UK: Cambridge University Press.

Zeki, S. (1993). *A vision of the brain.* Boston: Blackwell Scientific Publications.

Zeki, S., Watson, J. D. G., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. J. (1991). A direct demonstration of functional specialization in human visual cortex. *Journal of Neuroscience, 11,* 641–649.

Zihl, J. Cramon, D. von, & Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. *Brain, 106,* 313–340.

Zihl, J. Cramon, D. von, Mai, N., & Schmid, C. H. (1991). Disturbance of movement vision after bilateral posterior brain damage: Further evidence and follow up observations. *Brain, 114,* 2235–2252.

Zurek, W. H. (Ed.). (1989). *Complexity, entropy and the physics of information.* New York: Addison-Wesley Publishing Co.

Acknowledgements

For helpful comments on previous drafts I thank Mike Braunstein, Mike D'Zmura, Michael Gokhale, Perry Hoberman, David Hoffman, Julie Kwak, John Pyles, Barron Ramos, Whitman Richards, Kim Romney, Jay

Saul, Ian Scofield, Carol Skrenes, and Ken Vaden. The mathematical formulation of conscious realism was developed in collaboration with Bruce Bennett and Chetan Prakash. To both I am most grateful. Sole responsibility for any errors here is, of course, mine. This research was supported in part by a grant from the US National Science Foundation.

Figure Captions

Figure 1. The subjective Necker cube (Bradley & Petry 1977).
Figure 2. An illustration of the definition of a conscious observer.

Figure 1.

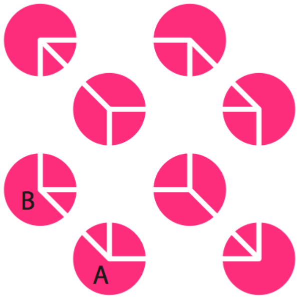

Figure 2.