

Sensory Experiences as Cryptic Symbols of a Multimodal User Interface

Donald D. Hoffman

Department of Cognitive Science

University of California

Irvine, California 92697 USA

ddhoff@uci.edu

Abstract

According to current theories of perception, our visual experiences match or approximate the true properties of an objective physical world. Ecological optics claims this match is direct, unmediated by psychological or computational processes. Computational theories claim that this match is indirect, the result of sophisticated algorithms that instantiate, e.g., the inferential methods of Bayesian estimation. The assumption of all of these theories that there is an objective, i.e., mind independent, physical world has proved incapable, so far, of yielding a scientific theory for the mind-body problem, a scientific theory of the relationship between conscious experiences and the brain. Therefore I explore, instead, the possibility that sensory experiences constitute a multimodal user interface (MUI) between the perceiver and an objective world, an interface useful precisely because it does not match, approximate, or resemble that world. I also explore *conscious realism*, the ontological claim that the objective world itself is comprised entirely of conscious agents. Together, MUI theory and conscious realism provide a straightforward solution to the mind-body problem, and entail *epiphysicalism*: physical objects, such as quarks and brains and stars, are constructed by conscious agents, but such physical objects have no causal powers.

Introduction

How does human vision work? What is the relationship between our visual experiences and the external world? The typical answer of non-specialists is that vision works like a camera, objectively reporting the shapes, colors, textures, motions, and lighting of the objective physical world. Our visual experiences, like the images from a camera, depict this physical world with high fidelity – except perhaps for the occasional optical illusion that usually occurs only in a psychologist’s laboratory, and simply because the eye has been deliberately tricked.

Current theories of vision agree with the non-specialist’s claim that our visual experiences depict the objective physical world with high fidelity, but disagree with their claim that vision achieves this result by working like a camera. According to these theories, it is true that the *eye* operates, to a first approximation, like a camera, with a lens that focuses an image on its retina; but the eye is just the starting point of vision. The retina of the eye has hundreds of millions of neurons, and their outputs are sent to the visual areas of the brain, which together engage roughly fifty billion neurons and twelve trillion synapses. In this regard, vision is not at all like a camera.

While current theories agree that vision depicts the physical world with high fidelity, these theories disagree about how precisely vision achieves this feat. Ecological theories, like that of Gibson (1987), claim that vision works by *directly* picking up properties of the physical world that are useful to the viewer, properties such as shapes, depths, motions, and colors. Just as one, and only one, string on a piano will resonate if a tuning fork is hit, so also the visual system resonates directly to useful properties of the physical world.

Computational theories, the modern descendants of the theory of unconscious inference of Helmholtz (1910/1962) and Alhazen (956–1039/Sabra, 1978), claim instead that vision infers properties of the physical world by means of sophisticated computations (Marr, 1982). Some claim that these computations are best described by regularization theory (Poggio, Torre, Koch, 1985): Vision solves the ill-posed problem of inferring properties of the physical world from the impoverished data of retinal images, and does so by instantiating in its computations certain regularizing principles that transform the ill-posed problem into a solvable well-posed problem. Others claim that these

computations are best described as instances of Bayesian estimation (Knill and Richards, 1996): Vision infers properties of the physical world as posterior distributions computed from three sources: image data, prior assumptions about the physical world, and likelihood functions that model the image-formation process. Computational theories often try to show that neural-network models of their computations could plausibly be instantiated in visual physiology.

Although the theories just described differ markedly from each other and from the typical non-specialist in how they claim vision works, what is most striking, for present purposes, is that they all agree with each other, and with the non-specialist, that our visual experiences, in the normal case, faithfully depict the objective physical world. Now I admit immediately that this assumption of faithful depiction is today almost universally accepted, that it has repeatedly proved useful in the study of vision and in all the sciences, and that to question it might seem, to some, simply ludicrous. But there are two problems that this assumption has, as yet, failed to resolve, or to give remotely plausible directions of inquiry, despite widespread and serious efforts: quantum measurement and the mind-body problem. It is, of course, premature to conclude that the assumption of faithful depiction of an objective physical world cannot, in principle, permit resolution of these problems. But it is not premature to consider other options. Here I ask if one can relax this assumption of faithful depiction, shared by all major current theories of vision, and still construct an empirically adequate theory. I begin by reviewing some properties of user interfaces.

User Interfaces

When you sit before your desktop PC, you have at your disposal a marvel of technology, a complex synthesis of sophisticated hardware and software. To fully understand your PC in all its complexity would require years of advanced training in fields such as physics, electrical engineering, circuit design, software engineering, operating systems, networks, and automata theory. If you simply want to type an email or create an illustration, you might reasonably be unwilling to subject yourself to such training. This is where the windows interface comes to your rescue. This user interface puts the power of the PC at

your disposal without requiring you to learn much about the PC. It hides the complexity of the PC behind colorful icons on your screen, icons that allow you to usefully interact with the PC and get your work done.

The power of a user interface is useful simplification (Schneiderman, 1998). Suppose, for instance, that you are writing a letter, and on your screen is a reddish square icon for the text file of that letter. Does this reddish square icon resemble that text file? Is the text file itself reddish or square? Of course not. Indeed the text file has no color or shape, but is instead a pattern of binary digits instantiated, e.g., as a pattern of voltages in certain memory circuits. Does the position of the reddish square icon on the screen tell you the position of the text file in the computer? If the icon is high on the screen does that mean it is near the top of the computer? Again, of course not.

Since the icon has color and shape, while the text file does not, and since the position of the icon does not indicate the position of the text file, then clearly the icon does not resemble the text file. Should we conclude that the icon is therefore a useless misrepresentation of the text file? Not at all. If you click on the icon, you can open the file and modify its contents. If you drag the icon to the recycle bin, you can delete the file. That is, the icon lets you interact with the text file, even though the icon in no way resembles that file. It is a cryptic symbol that effectively informs your behavior. So icons of an interface, in order to be useful, need not resemble the items they represent. Indeed, it is this very failure to resemble that makes an icon useful. We want to interact with colorful squares, not with myriads of voltages in memory circuits. An icon that displayed the myriads of voltages in their full complexity would, for purposes of getting a letter written, be useless. Faithful depiction would be, in this case, a nuisance and hinderance. The point of the interface is to let you interact easily with the computer while hiding its complexity.

Suppose now that you drag the reddish square icon to the recycle bin, and the text file is deleted. Is it the case that the movement of the icon to the recycle bin *caused* this deletion? No. The icon is a collection of colored pixels on the screen, as is the recycle bin, and neither the pixels of the icon, nor the pixels of the recycle bin, caused the deletion. Instead, the icon and recycle bin informed you how to move a mouse and how to click it. These motions and clicks of the mouse triggered a complex causal sequence of

states within the computer, which ultimately resulted in the deletion of the file. It might look otherwise to the casual user. After all, when the reddish square icon moves over the recycle bin, the lid of the recycle bin opens and the icon disappears. It is only natural to assume that the reddish icon caused the lid of the recycle bin to open, and that the recycle bin then destroyed the reddish icon and its text file. It is harmless, most of the time, for the casual user to assume that the windows icons have causal powers, or even to assume that the windows interface is itself all there is to the computer. Most of the time. We all know humorous stories of a young child or grandparent who makes this assumption, and tries to move the computer screen without the computer box, or asks why in the world is that ugly box attached to the screen. This confusion is a sign of a successful interface, one that provides such a simple and effective means of interacting with the computer that one can, for most purposes, simply take the interface to be identical to the computer. One can, with impunity, simply take the disk icon to *be* the hard drive, or the recycle icon to *be* the deleter of files. But of course this assumption, though convenient and usually harmless, is false. The icons are not the computer and they have no causal powers, appearances to the contrary notwithstanding. If the casual user's needs become more sophisticated than writing letters or making illustrations, they might then be forced to face what is, to them, a counterintuitive fact.

To summarize, a successful user interface has several characteristics. First, *friendly formatting*: the icons of the interface do not resemble what they represent; instead they drastically simplify what they represent, and do so in formats designed for ease of use, not for insight into the true natures of what they represent. Second, *concealed causality*: the icons of the interface have no causal powers; instead the interface as a whole hides, intentionally, the true complex web of causal sequences occurring within the computer. Third, *ostensible objectivity*: for many purposes one can act *as if* the interface were itself the total reality of the computer, and *as if* its icons were in fact causal agents; only for more sophisticated interactive needs do the limitations of this fiction become apparent and critical. Fourth, *clued conduct*: the icons effectively inform the actions of the user, allowing the user to interact with the computer even in total ignorance of the true nature of the computer.

Virtual Reality

Suppose that you and a friend go to a virtual-reality (VR) arcade and decide to play a game of virtual tennis. You each put on a helmet and body suit, and find yourselves on a grass court indistinguishable from the center court of Wimbledon. After a few warm up rallies, you start your first set. You serve the ball with heavy topspin and your friend smashes it down the line for a winner. The next few points go better for you, and soon you are engrossed in the game.

Behind the scenes, of course, is a supercomputer with many megabytes of software running the virtual tennis game. The visual and touch feedback you receive are a user interface to that supercomputer, an interface exhibiting the four characteristics, described in the last section, that are required for success.

First, it has friendly formatting. The tennis ball, for instance, informs you how to act in order to change the state of certain registers in the supercomputer so that you can win the point. It is much easier to track and hit a tennis ball than to learn the appropriate programming language and directly alter the registers in the computer. Indeed, even if one has learned the language and can directly alter the registers, one might not be able to do so as quickly and with as much likelihood of winning the point as the naïve user who simply tracks and hits the tennis ball.

Second, it has concealed causality. It looks, for instance, to the naïve user that it is contact of the racquet with the tennis ball that causes the ball to move with topspin. But the images of racquet and tennis ball are just images, without causal feedback to the computer. They inform your ground strokes, and these strokes are read by the computer and used to trigger a complex causal sequence of states, under program control, which lead to the incremental altering of the appropriate computer registers. These registers then cause new images of racquet and ball to be fed to the user. Causality appears to flow between ball and racquet, but instead it flows from the user's movements to the computer and thence, after a complex causal sequence, back to the images of ball and racquet.

Third, it has ostensible objectivity. For purposes of playing the game, it is convenient to forget the supercomputer and pretend that the virtual tennis court is the objective reality with which you are really interacting. For naïve players this is not simply a convenient fiction but, due to the effectiveness of the interface and their

ignorance of the supercomputer, their unquestioned assumption. Only for more sophisticated purposes, such as fixing a bug in the game, must one drop the fiction of objectivity, look behind the pretty interface, and deal with the computer itself.

Fourth, it has clued conduct. Although the racquet and ball have no causal powers, they do inform the player about how to run and stroke. Without the images of racquet and ball, the player's movements, though still causing changes in the state of the computer, could win a point only by sheer luck. But with those images and the clues they provide for running and stroking, the skill of the player becomes instead the determining factor.

We now apply these ideas about user interfaces and virtual reality to develop a theory of perception and conscious experiences.

MUI Theory

The MUI theory of consciousness claims the following: The conscious experiences of an agent are a multi-modal user interface (MUI) between that agent and an objective world, objective in the sense that its existence does not depend on that agent. MUI theory makes no claim about the nature of that objective world. In particular, it does not claim that the format or contents of the MUI in any way resemble or approximate that objective world. Conscious experiences of space and time, rocks and trees, colors and smells, might not resemble or approximate anything in that objective world. Thus MUI theory is more conservative than current theories of perception that claim that perception matches or approximates an objective physical world. MUI theory makes no such claim of matching or approximation. It leaves open the possibility that our sensory experiences, like the icons of user interfaces more generally, are effective precisely because they dumb down what they represent, and present it in a format designed not to give detailed insights but to facilitate useful actions.

This theory is perhaps most easily understood in the context of the virtual tennis game. In that game the player sees, holds and hits a tennis ball. If the underlying software is appropriately sophisticated, then the color, texture, feel, and three-dimensional shape of that tennis ball might be indistinguishable from a ball you could purchase at a sporting goods store. The naïve player might think that the tennis ball is objective, in the sense

that it does not depend on the player's perceptions for its existence. The player would, of course, be wrong. The tennis ball is entirely the player's construction, a construction triggered by sprays of photons from within the helmet. What exists independently of the player, in this example, is not the tennis ball but instead a supercomputer that in no way resembles the tennis ball. The computer is not round, yellow, fuzzy, and elastic; and yet, in this example, the computer is the objective reality the player truly interacts with whenever the player hits the ball. The same goes for the grass court, the net, the stands, and the racquets: the objective reality behind all of these is, in this example, a computer that resembles none of them.

A key step from this example to the MUI theory is this: Recognize that the computer in the VR example is itself just another icon, albeit a complex one, of our sensory interface. The consequence of taking this step is to realize that each one of your everyday sensory experiences – computers, rocks, brains, chairs and stars – is an icon of your user interface and, like the tennis ball of the VR example, depends on your perception for its very existence. Physical objects and properties are not the objective reality; they are simply elements of your user interface.

This might sound preposterous. Look, you say, this chair is not simply my sensory construction, a mere icon of my MUI. If I glance away, for instance, my friend still sees the chair, so it is there whether or not I look. And if I close my eyes, I can still touch the chair, so again it is there whether or not I look. So how can you say that the chair depends on my perception for its very existence?

These arguments are psychologically compelling but logically fallacious. Try them out in the setting of virtual tennis. Look, I say, this tennis ball is not my sensory construction. If I glance away, for instance, my friend still sees the tennis ball, so it is there whether or not I look. And if I close my eyes, I can still touch the tennis ball, so again it is there whether or not I look. Therefore the tennis ball does not depend on my perceptions for its very existence.

But these arguments are clearly false in the setting of virtual tennis. The only tennis ball to be had is the one that you experience. The objective reality, the computer, that triggers you to have a tennis ball experience does not itself resemble, in any way, a tennis ball. When you glance away, and your friend still sees the tennis ball, the ball they

see is the ball they construct, a ball dependent on their perception for its very existence. The ball your friend sees is not the same ball that you see, any more than if you have a headache and your friend has a headache, you are having the same headache. The tennis ball is not a “public object”: the ball you see is not numerically identical to the ball your friend sees. Nor is the grass court, net, or stands. There are no public objects in virtual tennis; nor are there any public objects in everyday life. The moon I see and the moon you see are not numerically identical. There is, of course, something that is objectively there, something that you and I both interact with, and that something triggers each of us to construct a moon. But that objective something need not resemble the moon any more than a supercomputer resembles a tennis ball.

It helps here to recall the well-known subjective Necker cube (Bradley and Petry, 1977), shown in Figure 1. In this figure you can see two cubes, one with corner A in front, call it cube A, and one with corner B in front, call it cube B; your perception switches from one to the other. It might happen that while you are seeing cube A, your friend might be seeing cube B; thus the cube that your friend sees is not numerically identical to the cube you see. It is in this same sense that the moon you see is not numerically identical to the moon your friend sees, or that anyone else sees. If you ask which cube is there in Figure 1 when you are not looking, the answer must be that there is no cube there when you are not looking. The cube you see only exists while you construct it. It is part of your MUI, not part of the objective world that does not depend on you for its existence.

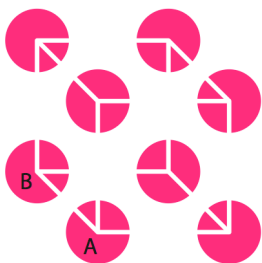


FIGURE 1. The subjective Necker cube (Bradley and Petry, 1977).

Still you might object. Look, you say, if that train hurtling down the tracks is a mere icon of your MUI, and depends on your perception for its very existence, then why don't you step in front of it? Then you'll find out for sure that it is more than just an icon.

But this is an elementary mistake. The reddish square icon on your computer screen does not resemble the text file it represents; there is not, inside the computer somewhere, a reddish square file. But you would be wise, nonetheless, not to drag that reddish square icon to the recycle bin, for then you might lose the text file and hours of your work. If the user interface is successful, then you must take its icons seriously, just not literally. So that train hurtling down the tracks is an icon of my MUI, an icon that I need to take seriously because my MUI is in good working order, but I would probably be mistaken to take it literally. The point of an interface is to dumb things down for ease of use. Whatever the objective reality is that triggers me to create the train icon, it is probably quite complex, and the train icon is probably a very dumbed-down symbol for it.

You might still object: Of course, you say, we know that the objective reality is more complex than a train. But we also know what that complex reality is. It is protons, neutrons, electrons, quarks and myriads of other particles whirling about in mostly empty space. So if that's the only difference between the train and objective reality that you're talking about, that's old news and not interesting news.

This is again an elementary mistake. It is like a person saying, Yes, I know the reddish square icon on my computer screen is not the objective reality of the text file. But I have right here my handy magnifying glass, and if you look real close through it you can see that the icon on the screen is really made of tiny little reddish pixels. There you have it; those reddish pixels are the true objective reality.

Of course those pixels are not, in that example, the objective reality, but are still part of the interface, still on the screen and not in the computer. In the same way, the protons, neutrons and other elementary particles are not the objective reality, but still part of the interface, albeit a part only encountered on close inspection. This assertion is, by the way, the standard (Copenhagen) interpretation of quantum theory (Wheeler & Zurek, 1983). On this interpretation the elementary particles have no dynamical properties unless

they are observed, and only when they are observed. That is, they are dependent on the observer for their existence. Thus, they are not part of the objective reality, but part of the user interface, the MUI. The reason, for instance, that the spin of an electron along an axis is “up” or “down” only if it is observed is the same reason that the image in Figure 1 is cube A or cube B only if it is observed: the spins, like the cubes, are constructions of the observer and therefore hold only during the act of observation.

Conscious Realism

Conscious realism claims the following: The objective world, the world whose existence does not depend on the perceptions of a particular conscious agent, consists entirely of conscious agents. According to this claim, consciousness is fundamental in the universe. It is not a latecomer that appeared by chance after billions of years of unconscious evolution.

Perhaps the easiest way to clarify the claim of conscious realism is to contrast it with other positions. To begin, conscious realism is not panpsychism nor entails panpsychism. Panpsychism is the claim that all objects – rocks, trees, atoms, stars – are themselves conscious (Hartshorne, 1937/1968; Whitehead, 1929/1979). Some versions of panpsychism restrict the class of objects that are conscious; atoms and trees, perhaps, but not rocks and chairs (Griffin, 1998). Conscious realism, by contrast, does not claim that any such objects are conscious. In the context of MUI theory, all such objects are icons within the user interface of a conscious agent, but that does not entail the claim that the objects themselves are conscious. To compare with the windows interface example, to claim that the objective reality behind a reddish square icon is a supercomputer is distinct from the claim that the reddish square icon is, itself, a supercomputer. This latter claim is obviously false, whereas the first is clearly true. Similarly, to claim that a tree, for instance, is an icon in the MUI of a conscious agent is distinct from the claim that the tree itself is conscious or a conscious agent. Conscious realism and MUI theory together entail that it is true that the tree is an icon in the user interface of a conscious agent, and false that the tree itself is conscious or a conscious agent.

Conscious realism is distinct from the idealism of Kant, in which he claims that the objective, “noumenal” reality, the “thing in itself” behind the phenomenal world of

appearances, cannot be described (Kant 1781/2003). Conscious realism and MUI theory together allow that the objective reality, the realm of conscious agents, can be described with complete mathematical rigor, so that scientific theories are possible. The appendix gives an example of a mathematical definition of a conscious observer; this, and related definitions, allow one to construct a dynamics of conscious agents that includes quantum theory as a special case (Bennett et al., 1989). From these remarks it should be clear that MUI theory and conscious realism are fully compatible with scientific naturalism, viz., the construction of rigorous mathematical theories with empirically testable predictions.

Conscious realism contradicts any form of physicalism, the doctrine that the objective reality consists only of spacetime, matter, energy, fields, and other elementary properties within the subject matter of physics. Conscious realism and MUI theory place all such physical properties squarely within the user interface, not in the objective reality. If conscious realism and MUI theory are correct, then physicalism suffers from a reification fallacy: it fallaciously identifies certain elements of the user interface with the objective reality behind the interface.

Conscious realism is a monism, not a dualism. Substance dualism, as advocated by Descartes, claims that the universe contains two kinds of substances, physical and mental, and that neither depends on the other for its existence (Descartes, 1641/1984; also Foster, 1991). Property dualisms typically claim that the objective reality is physical, but that in certain cases this physical reality can give rise to mental properties. If conscious realism and MUI theory are correct, both forms of dualism are false. All that exists are conscious agents and their experiences, i.e., the contents of their MUIs. What we call physical objects and properties are among the humbler contents of a MUI. But a MUI is itself an aspect of the structure of a conscious agent, not something separate, and therefore dual to, the conscious agent. To put it quite simply, when you look at a tree, the tree you experience is not in any way distinct from you, but is an aspect of you as a conscious agent. It is a symbol you construct as part of your own structure when you interact with some aspect of the objective world, an aspect which, although it triggers you to construct a tree, does not itself resemble a tree. You are not distinct from the tree; the tree is an aspect of you. Thus there is no dualism. It is helpful here to refer again to the subjective Necker cube of Figure 1. The cube you see in that figure only exists when you

construct it, and it exists only in your experience. That is, it is not distinct from you, but is instead an aspect of you. There is no dualism between you and the subjective Necker cube that you consciously experience.

We now apply MUI theory and conscious realism to resolve the mind-body problem.

The Mind-Body Problem

The formulation of the mind-body problem that is now almost universally accepted goes as follows: Assume that objective reality is physical, and explain how conscious experiences can arise from purely physical systems, e.g., systems like the brain. The problem with this formulation is that consciousness itself is not among the elementary physical properties, at least not in current physical theory, and it is not clear how to get consciousness to emerge from purely physical systems. That is, no scientific theory has yet been devised that can explain how consciousness could arise from purely physical systems. There are numerous physicalist “theories” of this mind-body problem, but even the authors of these theories know well enough that they are not scientific theories, just hints at directions where one might hope eventually to find a theory. A scientific theory makes, at a minimum, novel predictions with quantitative precision; it does not merely hint. The currently proposed hints suggest that conscious experiences might somehow arise from complexity (Edelman & Tononi, 2000), tracking relationships (Tye, 1995, 2000), information theory (Chalmers, 1995), functional architecture (Block & Fodor, 1972; Dennett, 1983), reentrant neural network dynamics and neural Darwinism (Edelman, 1992; 2004), quantum properties of microtubules (Penrose, 1994), and certain temporal patterns of neuronal firing (P.M. Churchland, 1995; P.S. Churchland, 2002, Crick & Koch, 1998; Koch, 2004). None of these hints has yet been turned into a theory that remotely approaches the minimal standards for a scientific theory. This is a remarkable failure, given that many brilliant minds have been working this problem for centuries.

This failure was the spark for the formulation of MUI theory and conscious realism. Perhaps the failure indicates that physicalism, which to its credit has successfully generated a wide range of good science, has reached a fundamental limit in its

explanatory power. Flat earth theories are great so long as you don't travel too far; locally, the earth is, indeed, approximately flat. But if you wish to circumnavigate the globe, you must abandon flat earth theory. Similarly, perhaps physicalism is fine so long as you are interested only in explaining structure and function of a MUI. But if you wish to go on to a scientific theory of consciousness, which includes the MUI as just one aspect, then you must abandon physicalism for a more adequate explanatory theory.

If MUI theory and conscious realism are assumed true, then the mind-body problem takes a new form. It is no longer the problem of bootstrapping consciousness from unconscious physical systems. It is instead the following problem: Assuming that the objective reality consists of conscious agents, how do “physical” objects – rocks, trees, stars, and atoms – arise from these conscious agents?

This mind-body problem has a straightforward answer. The “physical” objects are constructed by conscious agents in accordance with certain mathematically specifiable rules, rules that are now routinely being discovered and published by researchers studying human vision, computer vision, and other sensory systems. We now know, for instance, dozens of rules by which human vision constructs the colors, lightnesses, shapes, parts, and motions of objects (Hoffman, 1998). This mind-body problem is not a mystery for which we only have hints, it is an ongoing research enterprise in which our progress is sufficiently advanced that we build computer vision systems which functionally mimic the details of our scientific theories. Of course such computer vision systems are not themselves conscious, just as a mathematical theory of consciousness is not itself conscious, or as a computer simulation of weather will not get you wet. To expect more from such systems, theories and simulations is simply not to understand the scope, and limits, of science.

Causality and the Brain

Researchers studying the mind-body problem today almost universally assume that brain activity somehow causes, or is, conscious experience. Although, as we have discussed, they have failed yet to produce a scientific theory based on this assumption, they have several key arguments for believing it nonetheless. The arguments are as follows (e.g., Edelman, 2004). First, if you stimulate a person's brain that person often reports having

conscious experiences. Stimulate visual cortex, and they report visual experiences; stimulate somatosensory cortex and they report touch sensations. Second, if you damage a person's brain, that person often reports a reduced range of conscious experiences. Damage primary visual cortex and, if the damage is extensive enough, the person goes blind. Damage auditory cortex, and their hearing is impaired. Third, physiology and brain-imaging studies indicate that certain conscious experiences consistently correlate with certain patterns of brain activity. If a person sees a face, there is predictably a pattern of activation in their fusiform gyrus. If they imagine a face, the same areas of the brain are activated as when they see the face. Taken together, so the argument goes, these three classes of observations make a strong case that the brain causes conscious experience.

But these observed correlations do not, of course, *entail* that the brain causes conscious experience; to claim otherwise is to fall into a correlation/causation fallacy. And given that no one has yet been able to formulate a scientific theory of how brains could cause conscious experiences, it becomes quite plausible that the above-stated correlations indeed do not imply brain causation.

According to MUI theory and conscious realism the brain, being simply another icon in the MUI interface, has no causal powers. Just as the reddish square icon moving to the recycle bin does not cause the text file to be deleted, although that movement of the icon is perfectly correlated with such deletions, so also brain damage does not cause loss of conscious experience even if that brain damage is perfectly correlated with some such loss. Click on an icon and you can open the file, but the icon does not cause the file to open; similarly, stimulate the brain electrically and you can have a conscious experience, but the stimulation and subsequent brain activity do not cause the conscious experience. In short, because the brain is an icon of a well-functioning MUI, its behavior is well correlated with certain types of conscious experiences. But the brain never causes these experiences; it has no causal powers. Rather, the brain itself is a construction of a conscious agent, and exists as part of the MUI of that agent only for so long as the agent constructs it. Thus, according to MUI theory and conscious realism, the claim that the brain causes conscious experiences is simply an instance of the reification fallacy combined with the correlation/causation fallacy.

One position sometimes taken by physicalists toward the mind-body is *epiphenomenalism*, the claim that somehow conscious experiences arise, or are generated, from the brain but that these conscious experiences themselves have absolutely no causal powers (Jackson, 1982). This doctrine entails the problematic conclusion that my beliefs about my conscious experiences are not caused by those experiences, so that I could in principle have the same beliefs about my experiences whether or not I in fact had any experiences. This is close to a *reductio* of the doctrine. But the point of mentioning epiphenomenalism here is simply to contrast it to *epiphysicalism*, the position entailed by MUI theory and conscious realism. Epiphysicalism is the claim that physical objects and properties are constructed by conscious agents, but that these physical objects and properties themselves have absolutely no causal powers. They are elements of a MUI, and inform the choices and actions of conscious agents, but only the conscious agents have any causal powers. All the laws of physics, chemistry and biology describe regularities of a MUI, but not the true causal interactions of the objective world. We can often, with impunity, pretend that such laws describe the true causal interactions, just as we can often, with impunity, pretend that the windows interface is all there is to the computer. In our everyday life this pretense that our MUI is the objective world is a useful fiction, and possible to pull off only because of the quality of our MUI. It is when we face deeper questions in science that the fiction must be dropped if progress is to be made.

Evolution

An objection to MUI theory and conscious realism goes as follows. We know that the universe existed for billions of years before life emerged, and probably hundreds of millions of years longer before the first faint twinges of conscious experience emerged. The fossil record and modern genetic research make it clear that life has evolved along the lines outlined by Darwin. Clearly, so goes this objection, MUI theory and conscious realism are incompatible with modern evolutionary theory.

The response from MUI theory and conscious realism has several aspects. First, the mathematical models of modern evolutionary theory do not require a physicalist ontology. This mathematics works perfectly well as a model of interactions between

conscious agents. Indeed, evolutionary game theory (Maynard-Smith, 1982; Skyrms, 2000) is most naturally interpreted in just this way. There is simply no obstruction to having an evolutionary account of interacting conscious agents. So MUI theory and conscious realism are entirely compatible with a mathematically rigorous evolutionary theory.

Second, it is simply not true, according to MUI theory and conscious realism, that the universe existed billions of years before life and conscious experience emerged. That false world picture is the consequence of the reification fallacy, viz., mistaking a MUI for the objective world. Once one drops this fallacy, and recognizes that mathematical models of evolution can be applied to interacting conscious agents, then it should be possible to develop a new evolutionary theory which, when projected onto a MUI, gives the current physicalist evolutionary theory as a special case. The power of this new approach is that it holds promise that the same mathematical principles might ultimately underlie the dynamics of animate and inanimate systems. Within the current scientific paradigm, the evolutionary dynamics of life is distinct from the physical dynamics of inanimate systems. Within MUI theory and conscious realism this apparent dichotomy might dissolve, as a single mathematical framework handles both, with the dynamics of the inanimate appearing simply as a projection of a more complete dynamics onto certain aspects of a MUI, and the dynamics of life appearing as a projection of this more complete dynamics onto other aspects of a MUI.

Finally, evolutionary theory itself suggests that our conscious experiences are a MUI that does not resemble the objective world. Selection pressures do not evolve sensory systems toward exhaustive and faithful depictions of the objective world, but towards useful dumbed-down shortcuts that allow one to act quickly and appropriately. The race is to the swift, not to those encumbered with unnecessary complexity. Natural selection does not lead to truth, but to useful fictions for guiding swift actions. Humans, like roaches and lice, are another species among the tens of millions now on earth. We do not expect that the sensory systems of roaches give them special insights into the true nature of objective reality. Roaches have their niche to which their perceptions are adapted; humans have theirs. It would be a miracle, and a stroke of incredible luck, if the perceptions of either species were insights into the true nature of the objective world.

Conclusion

Some assumptions die hard. One can only imagine the initial reactions, centuries ago, to the early pronouncements that the earth is not flat and the stars are not near, that the earth is not the center of the universe and the sun does not revolve around the earth. The reaction was widespread incredulity. Any sane person could simply look around and see that these pronouncements were patently false. We now know that these sane people were simply suffering from the reification fallacy, mistaking aspects of the way things looked to them as true depictions of the objective world.

The assumption of physicalism, namely that the objective world consists of atoms, rocks, brains, and stars is, similarly, simply the consequence of the reification fallacy. But this physicalist assumption is very dear to us. We believe it from childhood without formal instruction. According to Piaget, as the child transitions from the sensori-motor stage to the preoperational stage, at about 24 months of age, it acquires “object permanence,” the knowledge that physical objects exist independent of the perceptions of the child. This illusion of object permanence, acquired in early childhood, stays with most of us the rest of our lives, and seems an inseparable component of our sense of reality. MUI theory and conscious realism ask us to give it up. This is much harder than giving up, say, flat earth; to give up flat earth we are just admitting that *one* of our perceptions is not the true objective reality. Now, MUI theory and conscious realism ask us to recognize that *all* of our perceptions are simply a MUI, they are not the objective world. This is a much bigger pill to swallow. But the failure of physicalism to give even one scientific theory of the mind-body problem may be the ailment that finally forces many to swallow it. It may go down easier with the recognition that MUI theory and conscious realism transform the mind-body problem from a mystery into a research enterprise in which substantial progress has already been made, while at the same time MUI theory and conscious realism incorporate all the methods and results of physicalist science by reinterpreting them as applying not to the objective world, but simply to a MUI.

Acknowledgements

For helpful discussions and comments on this paper I thank Mike Braunstein, Mike D’Zmura, David Hoffman, Julie Kwak, John Pyles, and Ken Vaden. This work was supported in part by a grant from the US National Science Foundation.

Appendix

One objection to MUI theory and conscious realism is the claim that they cannot be transformed from philosophical ideas into scientifically rigorous and predictive theories. They have, in fact, been so transformed (e.g., Bennett et al., 1989) and their mathematical formulation has proven powerful enough to derive quantum theory as a special case. It is, of course, beyond the scope of this paper to discuss this in detail, but one definition should give a flavor and help ease the objection. For details see, e.g., Bennett et al. 1989; 1991; Bennett, Hoffman, Kakarala, 1993; Bennett, Hoffman, Murthy, 1993; Bennett, Hoffman, Prakash, Richman, 1996.

Definition 1. An *instantaneous conscious observer* is a six-tuple, (X, Y, E, S, p, q) , where X and Y are measurable spaces, E is a measurable subset of X , S is a measurable subset of Y , $p: X \rightarrow Y$ is a measurable function satisfying $p(E) = S$, and q is a markovian kernel from Y to X .

The set X describes the formally possible conscious experiences of the observer, the subset E those conscious experiences that the observer can in fact construct, Y the sensory premises that formally could trigger the observer to construct a conscious experience, and S the sensory premises that can in fact trigger the observer to construct a conscious experience. The map p , often a projection, models the relationship between the formal structures of X and Y . The markovian kernel q models the construction of conscious experiences in E given sensory premises in S .

Given this definition of an instantaneous conscious observer, one can then go on to define a conscious observer that is not instantaneous, and a dynamics of such

conscious observers. The asymptotic behavior of the resulting dynamics can be shown to yield quantum dynamics as a special case (Bennett et al., 1989).

References

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 free: <http://www.cogsci.uci.edu/%7Eeddhoff/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., 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.

Block, N., Fodor, J. (1972). What psychological states are not. *Philosophical Review*, 81, 159–181.

Bradley, D.R., Petry, H.M. 1977. Organizational determinants of subjective contour: The subjective Necker cube. *American Journal of Psychology*, 90, 253–262.

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

Churchland, P.M. 1995. *The engine of reason, the seat of the soul: A philosophical journey into the brain*. Cambridge, MA: MIT Press.

Churchland, P.S. 2002. *Brain-wise: Studies in neurophilosophy*. Cambridge, MA: MIT Press.

Crick, F., Koch, C. (1998). Consciousness and neuroscience. *Cerebral Cortex*, 8, 97–107.

Descartes, R. 1641/1984. *The philosophical writings of Descartes*. Translated by J. Cottingham, R. Stoothoff, and D. Murdoch. Cambridge University Press.

Dennett, D. 1983. *Consciousness explained*. Boston: Little, Brown and Company.

Edelman, G. 1992. *Bright air, brilliant fire*. New York: Basic Books.

Edelman, G. 2004. *Wider than the sky: The phenomenal gift of consciousness*. Yale University Press.

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

Foster, J. 1991. *The immaterial self: A defence of the Cartesian dualist conception of the mind*. London: Routledge.

Gibson, J.J. 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.

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. 1998. *Visual intelligence: How we create what we see*. New York: W.W. Norton.

Jackson, F. 1982. Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127–136.

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

Knill, D., Richards, W. (Eds) 1996. *Perception as Bayesian inference*. Cambridge University Press.

Koch, C. 2004. *The quest for consciousness: A neurobiological approach*. Englewood, CO: Roberts and Company Publishers.

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

Maynard-Smith, J. 1982. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Penrose, R. 1994. *Shadows of the mind*. Oxford University Press.

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

Sabra, A.I. 1978. Sensation and inference in Alhazen's theory of visual perception. In Machamer, P.K. \& Turnbull, R.G. (Eds) *Studies in perception: Interrelations in the history of philosophy and science*. Columbus: Ohio State University Press.

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

Skyrms, B. 2000. Game Theory, Rationality and Evolution of the Social Contract. *Journal of Consciousness Studies*, 7, 1-2: 269–284.

Tye, M. 1996. *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.

Tye, M. 2000. *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.

Wheeler, J.A., Zurek, W.H. 1983. *Quantum theory and measurement*. Princeton University Press.

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