Published in *Philosophical Studies* (2007) 132: 301-330

Explanation and the hard problem[†]

Wayne Wright
wwright2@csulb.edu

ABSTRACT: This paper argues that the form of explanation at issue in the hard problem of consciousness is scientifically irrelevant, despite appearances to the contrary. In particular, it is argued that the 'sense of understanding' that plays a critical role in the form of explanation implicated in the hard problem provides neither a necessary nor a sufficient condition on satisfactory scientific explanation. Considerations of the actual tools and methods available to scientists are used to make the case against it being a necessary condition, and work by J.D. Trout that exploits psychological research on the hindsight and overconfidence biases is used to show that it is not a sufficient condition. It is argued, however, that certain intellectual and moral concerns give us good reason to still try to meet the hard problem's explanatory challenge, despite its extrascientific nature.

The hard problem of consciousness is to fully explain phenomenal consciousness—the subjective, qualitative dimension of our mental lives—in physicalistically respectable terms.[1] According to some, any explanation featuring only physical items, relations, and processes will always leave us wondering how and why there could be states with subjective, qualitative

---

[†] [Unfortunately, these Acknowledgments never made their way into the published article, despite the fact that I had included them in the corrected proof.] I am grateful to Carl Craver and Philip Robbins for many helpful comments on drafts of this paper, and for much lively discussion of the issues treated herein. Additionally, both have been gracious enough to allow me to cite work that is not yet published, for which I am thankful. Kent Johnson has been a beneficial source of questions and insights on the issues this paper tackles. David Chalmers' comments on an ancestor of this paper greatly aided my thinking about these issues.

[1] To avoid monotony, I will use a number of expressions to refer to phenomenal consciousness; e.g., 'consciousness', 'conscious experience'. Additionally, when I talk of a 'science of consciousness', I mean that to be one that includes both phenomenal and non-phenomenal aspects of consciousness.

properties; cf. Chalmers 1996, Jackson 1982, Levine 1983, McGinn 1991. Others deny that there is such a difficulty, including some who claim to have (at least the beginnings of) such explanations in hand; cf. Carruthers 2000, Dennett 1991, Dretske 1995, Tye 2000. My interest here is not to take a side in this debate. Rather, I aim to probe what is at stake in the debate, and I will argue that there is no scientific import to the hard problem as it is commonly understood. This result is significant at least because one might either take a failure to solve the hard problem to impose limits on what can be achieved in the scientific study of consciousness or have the idea that philosophical theories that meet the challenge of the hard problem thereby make some contribution to the scientific study of consciousness. My argument is straightforward:

(1) The hard problem is an explanatory problem.

(2) The kind of explanation central to the hard problem is not of a scientifically relevant sort.

(3) Therefore, the hard problem is an extrascientific matter.

By saying that the hard problem is extrascientific, I mean that it is a concern that is alleged to be of nontrivial consequence for the evaluation of scientific research, but that is generated by reasoning that ignores or distorts the tools, methods, and goals of (current and historical) actual scientific practice. This is not intended to be an exhaustive characterization of what it is for something to be extrascientific, but simply to capture the particular way in which this paper argues that the hard problem is extrascientific. A related issue that I will not pursue here is that the hard problem often serves as a stepping-off point—for philosophers of all manner of metaphysical stripes—for speculation about the essential nature of consciousness that goes well

beyond our best empirical evidence;[2] cf. Chalmers 2002 for an overview of the major metaphysical positions in the literature.

Since (1) is not at issue, what follows is solely directed at making a case for (2). Once (2) is in place, it follows that we should neither expect a solution to the hard problem from within science nor demand a solution to the hard problem as a necessary condition on constructing a fully adequate science of consciousness. Of course, we might still hope that scientific inquiry would be of help in addressing it.

1 The Intelligibility Requirement

Examining the explanatory challenge posed by the hard problem requires getting clear on the kind of explanation that is sought. The following does a good job of bringing out the relevant notion:

> For almost every natural phenomenon above the level of microscopic particles, there seems in principle to exist a reductive explanation; i.e., an explanation wholly in terms of simpler entities. In these cases, when we give an appropriate account of lower-level

---

[2] Relevant to this point is a comment from Francis Crick and Christof Koch about the recent burst of philosophical theorizing about consciousness:

> We shall not describe here the various opinions of philosophers, except to say that while philosophers have, in the past, raised interesting questions and pointed to possible conceptual confusions, they have had a very poor track record, historically, at arriving at valid scientific answers …. In recent years the amount of discussion about consciousness has reached absurd proportions compared to the amount of relevant experimentation. (Crick & Koch 1998/2003, pp.46-47; emphasis added)

> processes, an explanation of the higher-level phenomenon falls out … Once we have told the lower-level story in enough detail, any sense of mystery goes away: the phenomena that needed to be explained have been explained. (Chalmers 1996, p.42; my emphases)

Also consider this remark about explanation in terms of the low-level properties of physics:

> [In] showing how any instance of the phenomenon [to be explained] is itself implied by microphysical phenomena, we show that there is a sort of transparent epistemic connection between the microphysical and macrophysical phenomena. (Chalmers & Jackson 2001, p.351; my emphasis)

One more example, for good measure:

> The basic idea is that a reduction should explain what is reduced, and the way we tell whether this has been accomplished is to see whether the phenomenon to be reduced is epistemologically necessitated by the reducing phenomenon, that is, whether we can see why, given the facts cited in the reduction, things must be the way they seem on the surface. (Levine 1993/1997, p.548; my emphases)

The emphasized parts of these passages make plain that the sort of explanation at issue here involves a strong subjective component, along the lines of a 'sense of understanding' and an ability to 'see the fit' between the explanans and the explanandum. The thinking is that once we have the low-level physical details in place, the higher-level phenomenon appears to be a predictable consequence. Presumably, those who approach explanation this way would also include other, more objective, features in their account of explanation, but the subjective component clearly has a central role in determining whether or not a candidate explanation is satisfactory.

This conception of explanation as depending on a sense of understanding is consistent with much received opinion, and it tallies especially well with Hempel's (1965) inferential model of scientific explanation. On that account, explanations are arguments from explanatory facts to the expectation of the phenomenon to be explained. In the case of proposed explanations of phenomenal consciousness, the idea is that no matter how much detail about physical activity one packs in the explanation, a theoretically adequate description of phenomenal consciousness—one that captures the subjective, qualitative properties of experience—cannot be derived. If those properties are absent from the description, one has not truly explained consciousness. Call the requirement of the derivability of a theoretically adequate description the intelligibility requirement (IR).

The reliance on IR should raise our suspicions about the hard problem, as the inferential model was discarded long ago. Moreover, the inferential account was not offered to capture actual scientific practice, but rather as an account of how scientific explanation should proceed. Not only does the inferential model fare poorly in tasks that we would expect any acceptable account of explanation to handle (e.g., causal explanation; cf. Salmon 1990), it diverges so widely from actual scientific practice that it is normatively dubious. However, as Peter Carruthers (2004) points out, while IR is most closely associated with the inferential account, it is not altogether implausible to argue that something like IR must be met for scientific explanations fitting other models to be considered satisfactory.

One way of developing this line would be to argue that if an explanation does not make intelligible the relation between the putative explanatory facts and the target phenomenon, it is unclear just how our knowledge is advanced by that explanation. Thus there is no direct route from the failure of the inferential model to the extrascientific nature of the hard problem. For

example, Levine (1993/1997, p.549) explicitly rejects the Hempelian account of explanation while maintaining IR. Levine compares the explanation of why water boils that is achieved by the reduction of water to H_20 with what we could expect from any candidate physical or functional reduction of a conscious state. After sketching an account of the motion of H_20 molecules, intermolecular forces, vapor pressure and so forth, Levine (ibid.) concludes that it is "inconceivable that H_20 should not boil at 212°F at sea level." Returning to the longer passage from Levine quoted before, the idea is that there is an element of necessity to the explanation such that once one properly understands the explanans, it is impossible to coherently suppose that the explanandum could be otherwise; presumably this is to be attributed to the "seeing why" Levine mentions. If one can conceive of the explanans being in place but the explanandum being different or absent, then something has been left out of the explanation (ibid, p.548). Levine contends that this is precisely the case with consciousness, as the robustness of intuitions about zombies and inverted qualia indicate that no amount of physical or functional detail will prevent us from being able to coherently imagine those cases. Thus something significant about consciousness is lacking in the physical/functional story, which importantly is not the case with other phenomena for which we have satisfactory scientific explanations, such as the chemical theory of water.

Upon closer inspection, this prima facie response on behalf of a scientific conception of the hard problem falls apart. I have two main concerns. The first is that the demand that explanations of consciousness meet IR is likely rooted in a misplaced epistemic concern about the relation between an explanation's explanans and explanandum. The second is that IR-based conceptions of explanation rely on a sense of understanding that is an appalling failure as a guide to truth. The history of science is replete with cases of bad explanations being attended by a

profound sense of understanding, and empirical research reveals certain psychological biases that can account for much of the failure of the sense of understanding to guarantee that an explanation is one we should accept. Ultimately, I take intelligibility to provide neither a necessary nor a sufficient condition on satisfactory explanation, although it may very well be a desirable, but highly optional, characteristic. Thus I conclude that while we might be interested in whether the hard problem can be solved, the science of consciousness should not be counted on to—and should not be considered in any way deficient if there is a failure to—meet the challenge of the hard problem. At the end of the paper, once I have presented my case for why the hard problem is extrascientific, I will consider what sort of significance, if any, it might have.

At this early juncture, one might already be wondering about the relevance of psychological research on biases to a discussion of the hard problem. This use of psychological research can be justified straightforwardly by consideration of what it would take for the hard problem to pose a genuine challenge to the possibility of scientifically explaining consciousness. The friend of the hard problem might begin by insisting that the sense of understanding—which it seems hopeless to expect will ever attach to a candidate explanation of consciousness—should play a crucial role in scientific explanation. What support might this claim have? One nonstarter would be a direct appeal to the gripping phenomenology of the sense of understanding. The appeal to phenomenology alone simply does not intersect with scientific concerns and is ill-suited to establish a claim about the standards that need to be met by practicing scientists. Philosophers certainly have an important role to play in refining, clarifying, and criticizing various aspects of scientific practice, but such attempts promise to have an impact on science only when they actually engage with the sciences.

Although we need not take it as definitive of the thesis, Ronald Giere's (1999, p.5) characterization of some of the consequences of naturalism is helpful here: "any conclusions one reaches about the nature of science are subject to criticism based on theoretical, historical, psychological, or social investigations into particular scientific practices." By treating actual scientific practice itself as an object of inquiry and admitting of correction by many of the same sources that apply to scientific activity, naturalistic projects take on a number of virtues that make them favorable candidates for producing useful results regarding scientific activity and the limits of scientific knowledge. Of course, one is not barred from being more anti-naturalistic in one's ways and relying heavily on armchair reflection, but such a strategy places a gulf between philosophical and scientific activity; so much so that we should vigilantly inquire into the reasoning that is supposed to ground the relevance of the former to the latter. Pertinent to this point are the concerns raised by Stephen Stich and Jonathan Weinberg (2001) about our ability to use armchair reflection to "define the subject matter" in a way that, e.g., enables us to effectively judge the adequacy of scientific theories (cf. Jackson 1998); i.e., whether a candidate theory of X really is a theory of X or a theory of something else that is erroneously billed as a theory of X. Based on a review of relevant empirical research (which includes data from experiments they conducted), Stich and Weinberg argue that such a project cannot maintain its assumptions about the nature of concepts and the typicality of one's own intuitions.

A likely move to make at this point would be to attempt to buttress the appeal to epistemic phenomenology with examples drawn from the history of science in which a strong sense of understanding accompanied a good explanation. This is the way things have often gone in philosophical discussions of the form of explanation relevant to the hard problem; e.g., Levine's example of the chemical theory of water discussed above, Chalmers' (1996, p.42)

examples of earthquakes, heat, and lunar phases. Since it involves reflection on actual scientific activity, this appeal to the history of science could provide a suitable basis for demonstrating that the hard problem sets scientifically interesting limits on the study of consciousness.

A glaring weakness of the examples that have been cited in the literature, however, is that they have been selectively chosen from the available evidence, while examples that would lead us to question the link between "epistemic transparency" and good explanation have been neglected. As was noted above, there is no shortage of examples in the history of science of bad explanations that gave rise to the feeling that the target phenomenon was deeply understood. Moreover, there are also many cases in which the sense of understanding accompanied a good explanation, but the strong sense of "fit" led scientists to engage in questionable practice, such as dismissing competitors to their correct explanations for the wrong reasons; cf. Trout's (2002, p.227) discussion of Copernicus' "shuddering" at the thought of alternatives to heliocentrism. Consideration of the full track record of the sense of understanding should lead us to take a very careful look at claims that it has an important role in scientific explanation. Once we are aware of the sense of understanding's poor performance as a marker of good explanation, psychological research on biases immediately becomes relevant, as it has the potential to account for much of the tendency of both scientists and laypeople to accept explanations independently of whether they accurately capture the phenomena they are intended to. That sort of account would be quite valuable, since accepting an explanation that exhibits a high degree of fit between its explanans and explanandum but is not accurate when it comes to the features of the world it purports to explain would be detrimental to advancing scientific knowledge, as would rejecting an explanation that has a low degree of transparency but is accurate. Philosophical work that is supposed to make a contribution (even if only a negative one) to the sciences cannot insulate

itself from discoveries that might undermine its basic assumptions: in the case of the hard problem, experimental and historical findings that threaten the assumption that the sense of understanding is a core epistemic virtue of explanation. It is interesting to note that the selective use of evidence from the history of science in discussions of the hard problem looks to be a case of one of the biases that will be reviewed later in the paper: the confirmation bias.

Before proceeding, I should point out that the approach taken here has some obvious similarities to Patricia Churchland's (1997) critique of the role in sustaining the hard problem played by both ignorance and a failure of imagination. In particular, I share with Churchland a worry that an imprudent reliance on guesswork, gut feelings, and armchair reflection—all of which are susceptible to influence from limitations and peculiarities of human psychology—risks hobbling scientific progress by leading us to mischaracterize problem domains, creating unnecessary difficulty in recognizing the actual importance of scientific accomplishments, or seducing us into accepting scientific claims that are not as well-grounded as they intuitively seem. In the end, proper consideration of the limitations on and biases affecting our subjective powers of assessment should lead us to realize that the only genuinely promising way of evaluating the successes, failures, and boundaries of scientific inquiry is to intimately acquaint oneself with the tools, methods, and goals that figure in actual scientific practice.

2 Manipulation and explanation

To begin making my point that a scientific conception of the hard problem rests on an inapt epistemic anxiety about candidate explanations that fail to satisfy IR, I will sketch an account of scientific explanation inspired by the 'new mechanists'; cf. Bechtel 2002, Craver forthcoming, Glennan 2002, Machamer et al 2000. Although I am sympathetic to this approach, I am using it

here only as a means of exploring the mistaken motivations for IR-based accounts of scientific explanation. In other words, I will not argue for it, defend it from possible objections, or go into great detail about it, as the same conclusion could be reached in similar ways with other accounts of explanation that do not explicitly include anything like IR. The thrust of my argument in this section is that intelligibility is not a necessary element of satisfactory explanation, as scientists have at their disposal (and regularly make use of) tools and methods that both enable them to show a substantial connection between an explanation's explanans and its explanandum—thereby achieving certain vital scientific objectives that will be discussed shortly—and do not implicate the sense of understanding. It would not be far-fetched to claim that these explanations do (perhaps, must), in some way, render the target phenomenon intelligible. However, the sort of intelligibility that results is of a more practical sort, rather than the kind related to a subjective sense of understanding; i.e., 'seeing' the relationship between the explanans and explanandum, perhaps with an accompanying epistemic phenomenology. It is the latter, more robust form of intelligibility that is germane to the hard problem.

Consider an account on which explanation is intimately tied up with issues of control; cf. Woodward 2003. On this account, one has a satisfactory explanation of a phenomenon once one has identified a mechanism that enables manipulations and interventions that bring about non-accidental, fine-grained changes in the behavior of the phenomenon, or that allows us to answer a range of questions about how the phenomenon would behave if things were different. The second disjunct is especially relevant to systems that, for any number of reasons, we are actually unable to control; e.g., physical size, insufficiently developed technology, temporal distance. Along these lines, explanation goes a long way toward uncovering the structure of target phenomena in a way that advances the practical ends of scientific investigation. A great deal of

work will go into, e.g., tracking down a mechanism, decomposing it into its parts and their various relations and activities, and devising methods that enable one to hold some variables of the system steady while others are manipulated to bring about changes in the system's behavior. While IR might often be met by such explanations (e.g., "Oh, I see how such-and-such falls out"), there is no reason to think that IR must be met, or even will always be met. In describing the relevant mechanism, nothing about a 'theoretically adequate' description of the target phenomenon has to enter the picture. However, given the emphasis on manipulability, the explanatory facts themselves must be epistemically accessible in some way, and there must be a highly detailed account of the ways in which manipulations of the relevant factors would produce changes in the behavior of the target phenomenon. This more modest form of intelligibility is not enough to make the hard problem go away.

Returning to Carruthers' point mentioned earlier, friends of IR will claim that, even if an adequate description of the target phenomenon does not have to be derivable from a description of the proposed mechanism, we will only be satisfied with such an explanation when the descriptions of the relevant mechanism and the target phenomenon 'mesh'. This 'meshing' has to do with being able to understand that (or, perhaps, how) the cited mechanism as described is appropriately connected to the target phenomenon as described. However, this sort of meshing is not at all scientifically relevant, as the only apparent motivation for demanding it would be to assure us that the right phenomenon has been explained. Overlooked in such a condition on satisfactory explanation is that while there clearly is no codified set of standards by which scientists accept or reject explanations, the analytical and methodological resources at their disposal allow them to make those decisions without worrying about whether armchair reflection on a proposed explanation yields a sense of satisfactory understanding. By carefully analyzing

their experimental and observational data, scientists are quite capable of establishing a significant connection between a proposed mechanism and their target phenomenon, independently of any 'meshing'.

As a highly simplified example of this, suppose one argued that since scientists cannot directly observe another creature's experiences and we have not the faintest understanding of how it is that physical activity can give rise to conscious states, we have no guarantee that observable behavior will tell us anything about consciousness. If pressed, a scientist could provide much detail about her inference to the presence of experiences on the basis of behavioral data, background theory, etc. In the general case, the scientist begins by collecting data and discovers correlations between, e.g., a creature's behavior and the physical type of environment the creature was exposed to in the experiment. For example, there might be a correlation between the creature showing a marked preference for one piece of fruit over another when, holding all other variables constant, the first piece of fruit has a physical structure P (making it look red) and the second has a structure Q (making it look green). P and Q might not exhaust the possible physical realizations of red and green, so if the scientist is exploring whether a type of creature has color experiences of a particular sort, it would be natural to use multiple physical structures so that the only plausible commonality among the P_is was that they all appeared red, and similarly for the Q_is. If cautious, the scientist would apply a statistical technique to extract the latent variables of the data. These latent variables are high-level statistical constructs that statistically explain some of the variance or correlations present in the data. Further tests would help confirm the hypothesis that the statistical behavior of this construction is due to a 'hidden' property, such as a color experience, that is observable only by its effects on behavior.

These methods and techniques are standard equipment in disciplines that study complex systems and they are appropriate for inferring the existence of 'hidden' properties that are not directly measurable. The inference to others' experiences on the basis of behavioral data is just the sort of thing that the statistical methods of data analysis are designed to help us with. I am not suggesting that scientists need actually proceed this way. In actual practice, researchers may approach questions with strong background assumptions that make it easy to infer experiences from the data; e.g., with regard to how they handle verbal reports. The strategy outlined here would be particularly appropriate if there were no strong background assumptions. It is how scientists would characterize conscious experiences, were they called upon to justify their inferences. The availability of these tools and methods, and the challenges of studying the structure of complex systems are why causal models are built statistically, working with evidence such as 'P is correlated with B to degree r'. Correlations of zero or 1 are extremely rare, hence science is based on statistical generalizations, and not, e.g., first-order predicate logic. Thus scientific inquiry will not seek to tell us in any strong sense what phenomenal consciousness is, but rather will aim to investigate phenomenal consciousness and certain other aspects of the structure of our minds by discovering robust statistical regularities among, e.g., verbal reports, perceptual stimuli, saccading, blinking, skin galvanic response, gross bodily movement, brain activity, and (as theoretical posits) conscious experiences. Because they help us carve out relevant variables of the system under study and their relations to one another, those discoveries pave the way for better explanations of consciousness, manifested in increased abilities to manipulate, predict, retrodict, etc. conscious experience across a wide variety of circumstances.[3]

[3] The previous two paragraphs draw heavily on section 4 of Johnson & Wright (2006), in which it is

Explanations built on manipulation might be unsatisfying in that we cannot always 'see' the relation between a mechanism and the target phenomenon, but they are scientifically satisfying in that they meet scientists' (often primary) goal of bringing the phenomena they study under control. While certainly not at all guaranteed to deliver the kind of understanding of the target phenomenon that would satisfy IR, as was noted above, there is a more modest sense of intelligibility that comes with a good 'control' explanation. These explanations go well beyond mere simulations, as the mechanism figuring in a control explanation is what accounts for a potentially open-ended number of different simulations doing an equally good job of, e.g., predicting and retrodicting the behavior of the phenomenon. Such an achievement has considerable explanatory force, and it has an epistemic payoff far greater than, e.g., a systematized description of a large body of experimental results. Thus it is unclear what further 'mystery removal' scientists have to do. With a good control explanation in hand, scientists are equipped to meet various practical goals associated with their disciplines (e.g., correcting malfunctions, improving normal function, earlier and more precise diagnosis of malfunction) and to make strides forward in their investigation of phenomena related to the explained phenomenon.[4] Considering all the virtues that explanations can have without it, intelligibility of the kind connected to the hard problem seems ill-suited as a requirement on scientific explanation.

---

argued that to the extent that a theory of color is motivated by a desire to contribute something useful to scientific inquiry, the theory need not treat colors as anything more than what is represented by the latent variables of the statistical analyses of the relevant data.

[4] For more on the often primary emphasis placed on control in actual scientific practice, cf. Craver forthcoming.

In the light of the role statistical data analysis has played in the discussion so far, an opponent might respond by arguing that while population-based statistical explanations typically do not induce a feeling of understanding, this is simply an artifact of actual practice.[5] Certainly, such explanation does not by its very nature exclude the sense of understanding. More to the point, we should expect that breaking the macroscopic variance structure into its aggregate components of individual variation would facilitate a sense of understanding. The issue is one of the scale at which statistical explanations are routinely given, not one of floating free of the sense of understanding. In weighing this objection, the first thing to notice is that it is insufficient to establish that the sense of understanding is a necessary condition on explanation. The most it could hope to establish is that we should often expect to experience the sense of understanding when reflecting on components of a population-based statistical explanation. That is not the same as saying that we should accept such an explanation only if the sense of understanding sets in when considering those components. Additionally, the assertion about what we could or should expect is an empirical bet that would need to be backed by evidence.

The claim needed for the argument offered here is not that there is a natural opposition between the sense of understanding and explanations grounded in statistical data analysis. Rather it is that the value of the statistical techniques that researchers routinely make use of (e.g., principal component analysis, cluster analysis, multidimensional scaling) does not depend on anything like the sense of understanding. They are formal procedures for detecting or extracting

---

[5] I would like to thank an anonymous referee of this journal for suggesting that I consider this potential objection, as it neatly encompasses many of the responses I have encountered in discussing these issues across different circumstances.

some of the structure from researchers' data, or for simply organizing the data in various ways.[6] Thus there is nothing about these techniques themselves that creates an opening for an argument justifying the inclusion of epistemic phenomenology among the criteria that must be met by explanations developed on the basis of their use. A further elaboration of the potential objection just considered might be that, while the techniques themselves do not implicate the sense of understanding, researchers might have to rely on the sense of understanding in evaluating the products of their analyses. So, if the results of an analysis "don't look right" (e.g., not only are the results unexpected, but they resist attempts to "see a fit": How could that account for so much of the variation in the data?), the researcher might think that something has gone wrong in the design or in the way the data was handled, and make appropriate adjustments. It is tempting to focus on cases in which the analysis did not seem to fit and further examination revealed, for example, a confound, an experimental artifact, or an error in calculation or transcription, with subsequent improved analysis yielding a better fit. Echoing the observation made earlier about the selective use of examples from the history of science, exclusive focus on examples of this kind would neglect cases in which thorough investigation revealed no such features and repeated analysis (likely subjected to great scrutiny) continued to produce results that failed to tally with a sense of understanding. In such a circumstance, a determined insistence that the absence of

[6] Cf. Jolliffe 2002, Morrison 1976, Myers & Well 2003. Of course, there often are non-mathematical reasons for how certain issues are addressed, such as selecting a certain subspace for factor rotation to facilitate easier interpretation; cf. Abidi 2003, Meehl 1991, Thurstone 1947. Cf. the next footnote.

epistemic phenomenology tells us that the analysis is somehow deficient would likely be subject to a critique based on research on the overconfidence bias, which will be discussed shortly.[7]

To be clear, I am not implying that in the absence of IR, explanations will be easy to come by for any phenomenon, including consciousness. While I see no obstacle in principle to providing control explanations for aspects of phenomenal consciousness, I will not speculate about whether we should expect them to be achieved. It may turn out, though, that we will have to content ourselves with sketches of control explanations for various aspects of consciousness (as well as any number of other phenomena), in which homunculi rather than actual mechanisms are the main players. I will have more to say about this in section 4.

3 Sense of Understanding

Implicit in much of the discussion of the hard problem is the belief that our explanatory obligations would be discharged once we achieve an explanation that allows us to 'see' how physical processes could give rise to conscious states. Similar reasoning appears to be present in many philosophers' and scientists' approaches to explaining other phenomena: there is a tendency to suppose that if we are satisfied by our understanding of an explanation, that is a

---

[7] Not only is there a very real possibility that the sense of understanding will not accompany reflection on the account of the variation in the data provided by the extracted components, merely interpreting the components is not guaranteed to be easy. As Jolliffe (2002, p.64) points out, "It must be emphasized that although in many examples the PCs [principal components] can be readily interpreted, this is by no means universally true. There is no reason, a priori, why a mathematically derived linear function of the original variables (which is what PCs are) should have a simple interpretation." Jolliffe does also note that the first few principal components of an analysis quite often are easily interpretable.

signal that it is a good explanation. J.D. Trout (2002) has used empirical research on the hindsight and overconfidence biases to argue that intelligibility is not a useful gauge of an explanation's aptness, and thus we should rid ourselves of this tendency. I find Trout's argument compelling, and in this section I will review and add to the psychological research he cites in making his case. While Trout's concern is with scientific explanation in general, I want to focus here on the specific point that the research on these biases shows that the sense of understanding figuring in the hard problem is scientifically irrelevant. Given the poor track record of the sense of understanding that is sought, one might be led to wonder just what of substance is at stake in the hard problem.[8]

Trout 2002 argues that the sense of understanding—which is closely tied to the 'meshing', 'falling out', and 'mystery removal' mentioned earlier—does not permit us to recognize good explanations in its presence or bad explanations in its absence. The history of science is chockfull of instances in which an overwhelming sense of understanding accompanied a false explanation. Of course, it is not as though the presence of the sense of understanding indicates that an explanation is somehow deficient, as there are many examples of good explanations that have been attended by a strong sense of understanding. The truth or falsity of our explanations, and hence a major component of what determines their quality, simply is not revealed to us through the sort of 'epistemic phenomenology' that often accompanies 'seeing' how an explanation works. The sense of understanding is intertwined with a conviction that the

---

[8] Although I will not pursue the matter here, a related issue is that a number of studies call into question the reliability of our intuitions about our own minds, which would also seem to cast doubt on the relevance of (not) being able to 'see' how consciousness 'falls out' of some set of explanatory facts to whether an explanation is good or bad; cf. Maier 1931, Nisbett & Wilson 1977.

target phenomenon is an inevitable consequence of the explanatory facts. In a great many cases, this conviction feeds the impression that once understanding is achieved, there is no more explanatory work to do. Consider Ptolemy, who is an oft-cited example of someone who had a profound sense of understanding regarding a false theory. Much of the strength of his sense of understanding of his false theory is attributable to how well it cohered with rest of his body of beliefs, which was incomplete and riddled with falsehoods. Even worse, the sense of understanding led Ptolemy to ignore alternatives to his geocentric astronomical explanations and to refrain from pursuing deeper investigation into the principles governing the supposed motion of all objects toward the earth (cf. Trout 2002, pp.223-224). As difficulties were presented for his theory, Ptolemy would revise the theory, adding ever-more complicated details, rather than scrapping it altogether.

Relevant to this issue is an observation made by Francis Bacon, writing almost fifteen hundred years after Ptolemy, in a frequently quoted passage:

> The human understanding when it has once adopted an opinion (either as being the received opinion or as being agreeable to itself) draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects or despises, or else by some distinction sets aside and rejects; in order that by this great and pernicious predetermination the authority of its former conclusions may remain inviolate (Bacon 1620/1937, p.29)

Bacon's target is the selective use of evidence and counterevidence when it comes to evaluating beliefs that are strongly held. This is known today as the confirmation bias; cf. Nickerson 1998. The confirmation bias is just one well-researched example of the kind of errors that can be made in both scientific and everyday reasoning by relying too heavily on our subjective assessments.

On the one hand, we might come to strongly hold a false belief due to subjective biases; e.g., those that will be discussed in a moment. On the other hand, once that strongly held false belief is in place, it might be terribly difficult to shake because of other subjective biases; e.g., the confirmation bias. In the light of the impressive body of empirical evidence that calls into question our subjective powers of belief formation and preservation, it is puzzling that objective measures of explanatory adequacy would be put on a par with, and sometimes even made subordinate to, subjective measures.

What would account for the powerful intuition (admittedly felt by many) that understanding somehow clues us into whether or not an explanation is good? While there are likely several contributing factors, the hindsight and overconfidence biases are clearly relevant. The hindsight bias, also known as the 'knew it all along' effect, is revealed by a comparison of subjects' foresight judgments (e.g., estimations of the likelihood of possible outcomes without knowledge of how things actually turned out) and hindsight judgments (e.g., estimations of the likelihood of possible outcomes while informed by feedback about the actual outcome of the event); cf. Fischhoff & Beyth 1975, Dawes 1999, Hoffrage et al 2000; Schacter 2001, pp.145-149. Subjects systematically assign significantly higher probabilities to the actual outcome in the hindsight condition than they do in the foresight condition; i.e., informed retrospection leads subjects to assign themselves greater predictive acumen than is borne out by their actual predictive performance. In retrospect, the actual outcome looks to be strongly predictable from the antecedent facts, but forward-looking judgments reveal the actual outcome to really be highly unpredictable. Since explanation is formed retrospectively and is strongly sensitive to knowledge of the results of experimentation and observation, it is susceptible to a hindsight bias, one in

which one cannot help but see (in retrospect) that some phenomenon follows with necessity from the putative explanatory facts.

Daniel Schacter describes some of the more troublesome consequences of the hindsight bias:

> [If] we feel that we knew all along what would happen, then we may be less inclined to profit from the lessons a particular event or incident can teach us. But at the same time, the comforting sense that we always knew the way things would turn out makes us feel good about ourselves, inflating estimates of our own wisdom and prescience. (Schacter 2001, p.149).

With regard to its effects on subjective evaluations of explanations, a danger the hindsight bias leads to is that the perceived 'fit' between explanans and explanandum, and the sense that 'we knew it all along', inhibit pursuing further inquiry. This is particularly worrying because the sense of understanding that halts investigation might issue from nothing more than the coincidental coherence of false background beliefs (or an incomplete set of background beliefs) with experimental and observational findings.

Robyn Dawes' discussion of the crash of Western Airlines flight 903 in 1979 brings out the pernicious effects of relying on subjective assessments ('storytelling', in his words) rather than proper statistical analysis, in retrospective reasoning; cf. Dawes 1999, pp. 35-37 & 2001, pp.116-121. In the event in question, there is a single outcome, the crash, and five identifiable antecedents: fatigue, communication failure, vague communication, stress, and bad weather. There is no doubt that the combination of these five factors played a role in the crash. Certainly, it is highly unlikely that any one of those factors alone would have caused the crash. Retrospectively, we seem to have come to an explanation of the crash in terms of the

combination of the cited factors. However, in a forward-looking sense, we have no explanation at all, as:

- The cited factors routinely happen as a part of safe landings.
- Even if the cited factors had not been present, the crash could have followed other antecedents.
- Any number of other factors could have been (but tragically were not) present that would have prevented the crash, even with all five cited factors present.

As Dawes puts it, "[if] we were to do a prospective study of how well these precursors—either singly or in combination—predict whether a crash will occur, our measure of predictability … would most probably indicate gross unpredictability" (2001, p.118; original emphasis). The crash of Western 903 might appear to be a natural and predictable consequence of the cited factors—the kind of thing that one might expect to follow from the antecedent circumstances as described—but that appearance is misleading for the reasons just adduced. The way in which actual outcomes seem to follow directly and unquestionably from antecedent factors in hindsight reflection often leads us to credit ourselves with having accomplished something that we in fact have not, with the unfortunate result that we are left oblivious to the fact that our investigation has stopped well short of adequate explanation of the system in question; e.g., developing a statistical model that enables greater predictive success by identifying general antecedent factors that are associated with airplane crashes (ibid & Dawes 1999, pp.38-39).

The overconfidence bias is highlighted by studies of subjects' assessments of the accuracy of their judgment. Both subjects making judgments about mundane matters (e.g., frequency of various causes of death) and specialists offering judgments about their fields of expertise (e.g., physics, economics) have been found to assign confidence ratings to their

judgments that are vastly out of synch with their actual performance; cf. Fischhoff et al 1977, Henrion & Fischhoff 1986, Lichtenstein et al 1982. A confidence rating is simply an expression of the subject's assessment of the 'odds' that her judgment is correct. This bias is robust and produces a significant disconnect between subjects' sense of understanding, which is the basis for their confidence ratings, and the correctness of their judgments. The connection between the sense of understanding and confidence ratings is especially clear in the studies described in Fischhoff et al 1977, and it also figures in some of the suggestions made in Henrion & Fischhoff 1986 for how assessments of uncertainty in physical constants can be improved. It is compelling to think that the failure of the sense of understanding to guarantee correct judgment can be extended to a failure to guarantee good explanations. Subjective judgment plays a critical role in accepting or rejecting explanations on the basis of the sense of understanding, and the research on overconfidence shows such judgment to be liable to involve a mismatch between confidence and accuracy. Since confidence and accuracy are not tied to one another, we might be led to reject (i.e., assign very low or nil confidence to) a good explanation because we lack a strong sense of understanding of it, or accept (i.e., assign very high confidence to) a bad explanation simply because we find the sense of understanding attached to it irresistible.

Henrion and Fischhoff cite two potential contributing factors to overconfidence that seem to be related to the confirmation bias:

- "… the difficulty of thinking of reasons why one's best guess is wrong. Supporting reasons typically come to mind more readily than contradicting ones" (Henrion & Fischhoff 1986, p.796).

- "… the unequal treatment of such confirming and disconfirming evidence as is discovered … [People] have a considerable tendency to 'explain away' events that are inconsistent with their prior beliefs" (ibid).

Rabin and Schrag (1999) offer an extended argument for the claim that the confirmation bias induces overconfidence. In fact, the two biases would seem to feed each other, as selective use of evidence would tend to reinforce one's confidence in one's beliefs, and the more confident one is in one's beliefs, the more sensitive one will be to supporting evidence and the more determined one will be to discount or ignore troublesome evidence. As might be expected, based on its close connection with the confirmation bias, the overconfidence bias threatens to render us incapable of either exposing a false explanation or adding much-needed detail or refinement to a promising-but-incomplete explanation.

The hard problem is generated by a difficulty in understanding just how it is that physical processes could produce (realize, cause, be identical to, etc) states with phenomenal character. As the earlier quotes from Chalmers, Jackson, and Levine make clear, the kind of understanding sought involves more than recognizing the (more-or-less) objective virtues of a candidate explanation; e.g., predictive and retrodictive power, enabling control, providing a basis for further productive inquiry. Similar quotes abound in the consciousness literature. A strongly subjective element, discussed in metaphorical terms of 'meshing', 'seeing', 'falling out' and so forth, is thought to be a crucial part of adequate explanation. However, empirical findings such as the biases noted here should undermine our confidence in our subjective powers both of discerning good explanations from bad when deciding whether to accept an explanation, and of weeding out bad explanations after having accepted them. In the end, the sense of understanding sought in the hard problem may provide some comfort in our attempts to make sense of how our

best (but likely false or partial) theories about the nature and structure of the world could all be true at the same time, but it is not a helpful guide in our efforts to discover the actual nature and structure of the world. Even if one could 'solve' the hard problem by offering an explanation that allows us to 'see' the relation between physical processes and phenomenal states, that would offer no assurance that the proposed explanation is one we should actually embrace. Still missing would be a demonstration that the candidate explanatory facts provide information that can be put to use in, e.g., controlling the target phenomenon and (especially in cases in which manipulation is not actually possible) answering questions about different outcomes given various changes in the implicated factors; i.e., work along the lines of what was described in the previous section. As was argued in section 2, however, that work can be carried out in the absence of the sense of understanding.

4 Smoothing the rug and the real hard problem

Returning to a clarifying comment made in section 2, I do not mean to be suggesting that a scientific explanation of consciousness will be easy to achieve once we discard intelligibility as a component of explanation, nor am I in any position to speculate about whether such an explanation will be forthcoming any time soon, if ever. Additionally, the preceding discussion should not be taken as implying that there is no hard problem of consciousness at all, even if we were to dispense altogether with the standard way of understanding it. Human beings are complex systems that present serious methodological challenges to scientists in the best of circumstances. By any reasonable standard, consciousness would be counted among the human phenomena presenting the most difficult methodological challenges. The inability of scientists to directly observe another creature's conscious states, the well-documented problems with

accurate introspection and self-appraisal, the current state of neuroscience and its technologies, the differing interests and standards of even closely allied fields that study consciousness, and a whole host of other factors contribute to this situation. There certainly is a scientific hard problem of consciousness, but it is not hard for the reasons typically given in philosophical discussions of consciousness. It is vitally important to recognize what is actually difficult about explaining consciousness, at least as it relates to the aims of scientists, and to not allow our energy and focus to be misdirected, else we risk overlooking promising avenues for making headway on the task or being gulled into thinking progress is being made when in fact it is not.

Additionally, nothing that has been said so far evaporates or chips away at the mysteriousness of consciousness. It still seems just as puzzling that physical activity of any sort could produce a state with phenomenal character. It also still seems just as legitimate to be puzzled by this. The recognition that the hard problem that philosophers have been talking about is not the same hard problem confronting scientific investigation of consciousness does not (by itself, at least) establish that projects devoted to the former are confused or unjustified. In the light of the preceding discussion, however, it is fair to ask what payoff we might expect from research on the 'philosophical hard problem'.

I can think of two general kinds of benefits one might hope to reap from getting the sort of understanding of consciousness that philosophers have been seeking. The first was raised in passing at the end of the previous section: to smooth the rug. That is, to come up with some account of how our best ways of understanding the world could all be true at the same time, while also recognizing that our current understanding might be quite incomplete or simply mistaken. Something seems intellectually distasteful about having certain phenomena, especially something as prominent as consciousness, stand so far apart from the rest of our story about the

universe. We have a natural curiosity about how and why things happen as they do, and many philosophers throughout history have been concerned with making the fundamental structure of reality as rationally accessible as possible, or discovering what limits there might be on our knowledge. From this perspective, the project of developing an account of consciousness that makes intelligible its relation to physical activity is in keeping with many of the venerable aims of philosophy. Of course, assuming that such an account is achieved, it might have any number of results. It might reassure us that consciousness really is no different than any other complex phenomenon. It might prod us to think that there is something genuinely unusual about consciousness and that the universe is stranger than we already suppose it is. I am sure that there are many more possibilities and that philosophers would debate the consequences of any of them for how we should understand our place in the world.

The second potential benefit is closely related to the just-mentioned understanding of how we fit into the fabric of the natural world that might result from taming the hard problem. Consciousness is at the very core of our sense of who we are, both individually and as a species. The fact that it seems to be so hard to understand consciousness as a part of the natural order of things, especially when we do not appear to have any such difficulty when it comes to most other phenomena, seems deeply troubling, not just intellectually, but also morally. It has often been observed that the appearance of a chasm between us and nature—e.g., a sense that we are somehow 'special' and in some way not bound by the principles that govern the behavior of the world around us—can lead to all sorts of destructive (or otherwise undesirable) consequences in how we deal with our environments, non-human creatures, and even our own bodies and those of other humans. Perhaps a solution to the philosophical hard problem would bring with it a foundation for making progress in better handling these issues.

Robbins and Jack (submitted) have explored what they call 'the phenomenal stance' (i.e., treating another creature as a locus of phenomenal experience, akin to Dennett's 'intentional stance') and argued that part of taking up the phenomenal stance with respect to some creature is regarding it as a locus of moral consideration. Treating another creature as subject to phenomenal experience involves treating that creature as a proper object of moral concern, as considerations of pleasure and pain are at the core of moral sentiment. In fact, Robbins and Jack go beyond Benthamian considerations of pleasure and pain to include phenomenal experience in general; i.e., to regard a creature as subject to phenomenal experience of any sort is to thereby regard that creature as a subject of moral consideration. They offer a diagnosis of the psychological roots of the hard problem based on their notion of the phenomenal stance, arguing that there is cognitive neuroscientific evidence that shows a natural antagonism between the phenomenal stance and the physical stance. The physical stance is taken up when making "predictions based on the actual physical state of the physical object … worked out by applying whatever knowledge we have of the laws of nature" (Dennett 1971/1981, pp.4-5). If Robbins and Jack are right about the psychological difficulty of reconciling the phenomenal stance and the physical stance, and about the connection between the phenomenal stance and moral sentiment, then we have an explanation of the genesis of the hard problem that neither denies its reality nor makes consciousness out to be any more deeply mysterious than any other complex phenomenon.

We also, perhaps somewhat paradoxically, have a motivation for still wanting to solve the hard problem, despite it only being a psychological phenomenon with no obvious scientific or metaphysical consequences. Granting that there is great difficulty in integrating information across the domains of the phenomenal and the physical, recognizing the moral consequences of

our actions on the world around us may be problematical. There is no shortage of actual instances in everyday life of predicaments exhibiting this kind of struggle, and one does not have to believe that the environment itself deserves moral consideration to appreciate the potential for trouble here. The way we act on the physical environment and on the bodies of others considered 'merely' as bodies (rather than as closely tied to their phenomenal states) has the potential to, directly or indirectly, create great suffering or great pleasure. If we engage only the physical stance in reasoning about the environment and the bodies of other creatures, we are liable to overlook the moral consequences of our actions and risk either causing unintended harm or missing opportunities to better the lives of others.

Robbins and Jack are open to the possibility of future research decreasing the tension between the physical stance and the phenomenal stance, and there seems to be some good reason for optimism in this regard. Particularly relevant is the success of research on mental representation in lessening the appearance of incommensurability between the physical stance and the intentional stance, although by no means are all parties to the intentionality debates in agreement about the prospects for success of the project of naturalizing intentionality. It is safe to say, however, that the attempt to give an account of intentionality that makes clear that it is part of the natural order, is now nowhere near as widely believed to be hopeless as it once was. By forging a closer epistemic link between the physical and the phenomenal, perhaps a solution to the hard problem would reduce the antagonism between the phenomenal stance and the physical stance, allowing us to more easily think of consciousness as part of the natural order of things. We might hope that this would be accompanied by an improvement in our ability to acknowledge moral dimensions of our actions to which we are currently blind. Supposing that it can be shown that work on the philosophical hard problem could reasonably be counted on to

deliver such a morally advantageous result, the possibility of such a gain by itself would make the philosophical hard problem significant, despite its extrascientific nature.

5 References

Abidi, H. 2003. "Factor rotations." In M. Lewis-Beck, A. Bryman, T. Futing (eds.) Encyclopedia for research methods for the social sciences. Thousand Oaks (CA): Sage.

Bacon, F. 1620/1937. "Of Great Place," in C. W. Elliot (ed.) Essays, Civil and Moral - The Harvard Classics. New York: P. F. Collier and Son.

Bechtel, W. 2002. "Decomposing the brain: A long-term pursuit." Brain and Mind, 3, 229-242.

Carruthers, P. 2000. Phenomenal Consciousness. Cambridge: Cambridge University Press.

_____. 2004. "Reductive explanation and the explanatory gap." Canadian Journal of Philosophy, 34.

Chalmers, D. 1996. The Conscious Mind. New York: Oxford University Press.

_____. 2002. "Consciousness and its place in nature." In D. Chalmers (ed.) Philosophy of Mind: Classical and Contemporary Readings. New York: Oxford University Press.

Chalmers, D. & F. Jackson. 2001. "Conceptual Analysis and Reductive Explanation." Philosophical Review, 110, 315-360.

Churchland, P.S. 1997. "The Hornswoggle Problem." Journal of Consciousness Studies, 3, 402-408.

Craver, C. forthcoming. Explaining the Brain.

Crick, F. & C. Koch. 1998/2003. "Consciousness and Neuroscience." In B. Baars, W. Banks, & J. Newman (eds.) Essential Sources in the Scientific Study of Consciousness. Cambridge, MA: MIT Press.

Dawes, R. 1999. "A message from psychologists to economists: mere predictability doesn't matter like it should (without a good story appended to it)." Journal of Economic Behavior & Organization, 39, 29-40.

_____. 2001. Everyday Irrationality. Boulder, CO: Westview.

Dennett, D. 1971/1981. "Intentional Systems." In Brainstorms: Philosophical Essays on Mind and Psychology. Cambridge, MA: MIT Press.

_____. 1991. Consciousness Explained. Boston: Little, Brown.

Dretske, F. 1995. Naturalizing the Mind. Cambridge, MA: MIT Press.

Fischhoff, B. & R. Beyth. 1975. "'I knew it would happen': remembered probabilities of once-future things." Organizational Behavior and Human Performance, 13, 1-16.

Fischhoff, B., P. Slovic, & S. Lichtenstein. 1977. "Knowing with certainty: the appropriateness of extreme confidence." Journal of Experimental Psychology: Human Perception and Performance, 3, 552-564.

Giere, R. 1999. Science Without Laws. Chicago: University of Chicago Press.

Glennan, S. 2002. "Rethinking mechanistic explanation." Philosophy of Science, 69, S342-S353.

Hempel, C. 1965. Aspects of Scientific Explanation and Other Essays in the Philosophy of Science. New York: Free Press.

Henrion, M. & B. Fischhoff. 1986. "Assessing uncertainty in physical constants." American Journal of Physics, 54, 791-798.

Hoffrage, U., R. Hertwig, & G. Gigerenzer. 2000. "Hindsight bias: a by-product of knowledge updating?" Journal of Experimental Psychology: Learning, Memory, and Cognition, 26, 566-581.

Jackson, F. 1982. "Epiphenomenal Qualia." Philosophical Quarterly, 32, 127-136.

_____. 1998. From Metaphysics to Ethics: A Defense of Conceptual Analysis. Oxford: Oxford University Press.

Johnson, K. & W. Wright. 2006. "Colors as properties of the special sciences." Erkenntnis, 64, 139-168.

Jolliffe, I.T. 2002. Principal Component Analysis. New York: Springer.

Levine, J. 1983. "Materialism and Qualia: The Explanatory Gap." Pacific Philosophical Quarterly, 64, 354-361.

_____.1993/1997. "On Leaving Out What It's Like." In N. Block, O. Flanagan, G. Guzeldere (eds.) The Nature of Consciousness: Philosophical Debates. Cambridge, MA: MIT Press.

Lictenstein, S., B. Fischhoff, & L.D. Phillips. 1982. "Calibration of probabilities: State of the art to 1980." In D. Kahneman, P. Slovic, & A. Tversky (eds.) Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.

Machamer, P., L. Darden, & C. Craver. 2000. "Thinking about mechanisms." Philosophy of Science, 67, 1-25.

Maier, N. R. F. 1931. "Reasoning in humans: 2. The solution of a problem and its appearance in

consciousness." Journal of Comparative Psychology, 12, 181-194.

McGinn, C. 1991. The Problem of Consciousness. London: Blackwells.

Meehl, P. 1991. Preface to C.A. Anderson & K. Gunderson (eds.) Selected Philosophical and Methodological Papers. Minneapolis: University of Minnesota Press.

Morrison, D. 1976. Multivariate Statistical Methods. New York: McGraw-Hill.

Myers, J. & A. Well. 2003. Research Design and Statistical Analysis. Mahwah (NJ): Lawrence Erlbaum Associates.

Nickerson, R. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." Review of General Psychology, 2, 175-220.

Nisbett, R. & T. Wilson. 1977. "Telling more than we can know: Verbal reports on mental processes." Psychological Review, 84, 231-259.

Rabin, M. and J. Schrag. 1999. "First Impressions Matter: A Model of Confirmatory Bias." Quarterly Journal of Economics, 114, 37-82.

Robbins, P. and A. Jack (submitted). "The Phenomenal Stance."

Salmon, W. 1990. Four Decades of Explanation. Minneapolis: University of Minnesota Press.

Schacter, D. 2001. The Seven Sins of Memory. Boston: Houghton Mifflin.

Stich, S. & J. Weinberg. 2001. "Jackson's Empirical Assumptions." Philosophy and Phenomenological Research, 62 637-644.

Thurstone, L. 1947. Multiple Factor Analysis. Chicago: University of Chicago Press.

Trout, J.D. 2002. "Scientific explanation and the sense of understanding." Philosophy of Science, 69, 212-233.

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

Woodward, J. 2003. Making Things Happen: A Theory of Causal Explanation. Oxford: Oxford University Press.