

## **Why naturalize consciousness?<sup>†</sup>**

Wayne Wright

Dept. of Philosophy \* Cal State – Long Beach \* wwright2@csulb.edu

Abstract: This paper examines the relevance of philosophical work on consciousness to its scientific study. Of particular concern is the debate over whether consciousness can be naturalized, which is typically taken to have consequences for the prospects for its scientific investigation. It is not at all clear that philosophers of consciousness have properly identified and evaluated the assumptions about scientific activity made by both naturalization and anti-naturalization projects. I argue that there is good reason to think that some of the assumptions about physicalism and explanation made by the parties to the debate are open to serious doubt. Thus this paper is an invitation for those inquiring into whether consciousness can be naturalized to more carefully consider the expected payoff of such efforts.

### **1 Naturalizing projects**

It is widely accepted that phenomenally conscious states do not present themselves as brain states, information processing states, quantum mechanical states, or anything else that would easily fit within the ontology of one or another scientific field.<sup>1</sup> However, conscious experiences do not seem to be of a kind that is altogether discontinuous with the world around us. After all, I think that the twinge I feel in my knee is due to an old sports injury,

---

<sup>†</sup> Parts of this paper were presented at several different venues and I am grateful to those audiences for their comments and questions. Special thanks are due to Jeff Barrett, David Chalmers, Carl Craver, Kent Johnson, Penelope Maddy, Philip Robbins, and Charles Wallis for valuable feedback and discussions that greatly aided the development of this paper. Lastly, I would like to thank the anonymous referees of this journal for their most helpful reports.

<sup>1</sup> To avoid monotony, several expressions are used to refer to phenomenal consciousness; e.g., ‘consciousness’, ‘conscious experience’, ‘phenomenal experience’.

the coarseness I feel when I rub a piece of sandpaper between my thumb and forefinger fits with the rough texture of the sandpaper, and the vivid hues I am so dazzled by when gazing at the setting sun look like they are out there in the world. Despite there being some continuity between phenomenal states and the rest of the world, the great difficulty in understanding how they might fit entirely within the natural order is often taken to support anti-physicalist views. Such views can include a commitment to claims about the impossibility of a full scientific explanation of consciousness or a more far-reaching skepticism about the prospects for a legitimate scientific study of consciousness. It is at this point that naturalizing projects enter the picture. While there is no consensus about the specifics, the basic idea is that a naturalistic account of consciousness would show that the subjective, qualitative properties of certain mental states can be squared with the physicalistic worldview of the sciences.<sup>2</sup>

A commitment to physicalism as an essential part of the sciences can be found in remarks from parties on differing sides of the consciousness debates. Frank Jackson's much-discussed "Mary's Room" thought-experiment brings this out, unpacking Mary's "[knowledge of] all the physical facts about us and our environment" in terms of her knowing "everything in completed physics, chemistry, and neurophysiology" (Jackson 1986, p.291; emphasis in the original). William Robinson (2004) voices what I take to be relevant, widely held views about naturalism and physicalism:

[Supporting] materialism means showing how difficult questions can be brought within the purview of science. (p.42)

---

<sup>2</sup> This characterization is drawn from that given by Stich (1992) for naturalistic accounts of mental representation; cf. also Stich (1996) and Stich & Laurence (1996).

[Naturalism is the] commitment to recognizing no entities other than those recognized in the natural sciences. (p.53)

David Chalmers (1996, p.128) holds that “[sometimes] ‘naturalism’ is taken to be synonymous with ‘materialism’, but it seems to me that a commitment to a naturalistic understanding of the world can survive the failure of materialism.” More will be said later about Chalmers’ take on naturalism and physicalism, but for now we can just consider Emmett Holman’s (2006, p.252n21) reaction:

By ‘naturalizing’ intentionality is meant, of course ‘physicalizing’ it ... David Chalmers (1996) calls his [anti-physicalist] theory of mind ‘naturalistic dualism’, but this is not the kind of naturalism that naturalizers of intentionality have in mind.

A colorful expression of the connection between physicalism and the scientific outlook, one which quite forcefully reveals the expected payoff of naturalization projects, comes from Jerry Fodor (1990, pp.202-203; quoted in Stich & Laurence 1996, p.189n2):

If it turns out that the physicalization – naturalization – of intentional science .... is impossible, ... then it seems to me that what you ought to do is do your science in some other way ... If you really can’t give an account of the role of the intentional in the physical world ... by Christ ... we should stop spending the taxpayer’s money.

Adding some detail, being a naturalist about X requires a commitment to X being fully explicable in terms of entities (properties, etc.) that figure in some scientific field, and the thinking is that making good on that requires that a physicalistic account of X can be had. So, a physicalistic theory of consciousness would offer the reassurance that its scientific study can, in principle at least, deliver an account that is in on a par with the rest of the sciences in terms of its adequacy, completeness, etc. Likewise, a convincing argument

against (or a miserable run of failed arguments for) the possibility of physicalizing consciousness will have negative results for the prospects for its scientific study.

In what follows, central aspects of the philosophical debate over naturalizing consciousness are critically examined. The conclusion reached is that the debate is ill-founded and neither anti-physicalist arguments nor naturalizing projects can have the consequences for the scientific investigation of consciousness often claimed for them. To be clear, I am only engaging with those who see the battles over the physicalistic status of consciousness bearing on its scientific study; this leaves out those who think it is clear from the outset that a science of consciousness is impossible or whose study of consciousness is directed toward wholly extrascientific purposes. There are two main lines of argument. The first considers how questions of ontological legitimacy are routinely decided in the sciences. Traditional metaphysical arguments, for or against physicalism, are irrelevant to how such questions should get decided. The second argument is for the claim that worries about an “explanatory gap” (Levine 1983, 1993/1997) or a “hard problem” of explaining consciousness (Chalmers 1996) – which are a major motivator of the disputes about whether consciousness can be naturalized – turn on a defective model of scientific explanation.

While I am certainly not the first to doubt the scientific relevance of various philosophical concerns, more needs to be done to encourage philosophers of consciousness, specifically, to give careful consideration to the actual significance of their labor to scientific research. The pessimistic view developed here might plausibly be seen as the initial result of an application to consciousness of something in the spirit of Penelope Maddy’s (2007) *Second Philosophy*. Central to *Second Philosophy* is to regard naturalism as more of a methodological doctrine than a metaphysical one, and to evaluate methods, proposals, and

research topics on the basis of their promise to contribute to productive inquiry.<sup>3</sup> I see mainstream debates about the metaphysics of consciousness faring rather poorly on such an evaluation and there being a need for philosophers of consciousness to re-orient their work so that it is better positioned to yield genuinely fruitful results. The discussion of explanation in the latter part of the paper includes a small first step in that direction.

## **2 The legitimating relation**

Differing physicalist theories agree that everything either is physically “basic” or bears a “legitimating” relation to the basic entities; cf. Crook & Gillett (2001, p.336), Schaffer (2003, p.506). An obvious first step in fleshing out physicalism is to take the fundamental entities of physics as the basic entities of one’s physicalistic ontology. While there are well-known problems in specifying the basic entities of a physicalistic ontology, they will not be pursued here.<sup>4</sup> Waiving those worries won’t by itself clear the way for physicalism to be

---

<sup>3</sup> The Second Philosopher “is simply engaged in inquiry” (Maddy 2007, p.62) and “begins from commonsense perception and proceeds from there to systematic observation, active experimentation, theory formation and testing, working all the while to assess, correct, and improve her methods as she goes” (ibid., p.2).

<sup>4</sup> Seth Crook and Carl Gillett (2001) argue that while it is tempting to think of physicalism as a part of science, there are obstacles to it meeting the standards of a scientific hypothesis in good standing; cf. also Chomsky (1968, 2000), Gillett (2001), Hempel (1969), Schaffer (2003), and van Fraassen (1996, 2002). Cf. Melnyk (1997) and Poland (1994) for leading accounts of physicalism criticized in Crook & Gillett (2001). Putting Crook and Gillett’s position in simple form, a physicalism grounded in current physics is probably false and one grounded in a future physics is indeterminate at best and is likely trivial. Such qualities would eliminate physicalism as a genuine scientific hypothesis. They suggest that it is best

used in adjudicating questions of scientific import. There is still the matter of whether all scientifically respectable kinds must exhibit a certain relation to the fundamental physical entities. Of specific concern is the relevance of naturalizing projects to issues of scientific ontology, even if it is granted that the sciences have a commitment of some sort to physicalism. This section argues that the sciences need little or nothing from such projects in order to proceed appropriately. The idea that (quasi-) armchair reflection about metaphysical issues can affect the scientific legitimacy of some kind X gets things backwards. A kind's role in our best scientific theories justifies its inclusion in our scientific ontology and, if it were to turn out that physicalism is in some way scientifically relevant, we should infer that a kind is physicalistically acceptable from its scientific legitimacy, not the other way around. Thus naturalizing projects as typically conceived are either otiose or altogether misdirected.

---

conceived only as a framework in which to examine various metaphysical issues. While physics and other sciences might have a role in determining what counts as physical, physicalism does not guide scientific inquiry and "is not directly, or necessarily, connected to physics or its findings" (Crook & Gillett 2001, p.357). If physicalism is an extrascientific thesis, there is apparently no need for naturalizing responses to anti-physicalist arguments.

There is also the issue of our current ignorance about whether fundamental physical entities exist at all; cf. Suppes (1978, p.7). Jonathan Schaffer (2003) makes a compelling case for agnosticism about fundamental physical entities, based largely on the empirical nature of the debate and the lack of anything like a clear-cut basis for preferring one position to the other; cf. also Greene (1999, p.142). A consequence of this admission of ignorance is that, at present, proceeding as though nature admits of a tidy description in terms of basic building blocks and finitely storable laws relating them to everything else involves a nontrivial assumption. Also nontrivial is the idea that even if nature were so agreeable, we could make use of such a description were it given to us; cf. Teller (2001, pp.393-394).

Returning to the discussion of section 1, a naturalist might endorse physicalism as a necessary, but not sufficient, condition on scientific legitimacy. That is, if one demonstrates that X fits within the physicalistic worldview, all that one has secured is that X is a candidate for appearing in scientific theories; something else is needed to determine whether X is to be incorporated in our scientific ontology. Michael Tye (1992) offers a “physicalism plus” naturalism that can be used to focus the discussion of the general approach. On his account, for a kind to count as natural, it must:

1. “bear the same ontic relationship to lower level physical items as do the physical items quantified over and referred to in higher level physical laws generally” (ibid., p.434).
2. “participate in causal interactions describable in scientific laws” (ibid.).

These conditions are presented as individually necessary and jointly sufficient for X to qualify as natural. For Tye, the ontic relationship in (1) is realization, rather than type reduction (ibid, pp.436, 438).<sup>5</sup> At root, realization is a relation between a lower-level kind and a higher-level kind, such that at any time t the tokening of the lower-level kind nomologically necessitates the tokening of the higher-level kind, but the converse does not hold (ibid., p.436). The realization relation is expected to ultimately ground out in the microphysical realm, with any number of stops at intervening levels along the way.

---

<sup>5</sup> Tye is explicit that his appeal to realization is intended to avoid the difficulties plaguing stronger, reductive projects; cf. also Stich (1992). The concerns raised about this weaker doctrine in the following also apply to the stronger accounts. Tye’s (1994/1997, 1995, 2000) subsequent work on consciousness, which argues for a reduction of phenomenal character to a species of representational content in order to bring phenomenal consciousness within the domain of cognitive psychology, abandons the account of naturalism offered in his (1992).

An obvious reason why (2) is needed to establish that a kind is natural is that it is a trivial matter to cook up kinds that are physically realized (i.e., that satisfy (1)) but that are far from scientifically legitimate. For example, a glurg is a head of cabbage, or a one acre region containing at least one short ton of anthracite coal if and only if no living thing is currently passing through the Van Allen radiation belts, or a flounder with a minnow in its mouth. Glurg is certainly a physically realized kind, but it does not, nor should it be expected to, appear in any branch of science. The second condition gets us that glurg is not a natural kind, but that stratovolcanoes, hantaviruses, and white dwarf stars are. When it comes to establishing that consciousness is every bit as natural as these latter kinds, philosophers have the task of showing how (1) might be satisfied and it is left to scientists to make good on (2). A striking consequence of this conception of naturalism is that X's claims to naturalness could be undermined by arguments challenging its physical realization (or by continual failure to prove that X is physically realized), even if X appears in our best scientific theories.

While the place of (2) in an account of what it is for a kind to be natural might be apparent, less clear is the role of (1). One major reason given for including something like (1) in an account of naturalism is that without an adequate physicalism in place, there is no ruling out "non-natural" positions such as substance dualism; cf. Tye (1992, pp.432-433), Crook & Gillett (2001, p.345), P.S. Churchland (1993/1997, p.127). Tye provides a representative expression of this anxiety, protesting that relying only on a kind's appearance in scientific laws or theories to determine its naturalness leaves open that the kind in question is really the manifestation of some otherworldly "stuff." He writes of the use of (2) alone in determining whether mental states are natural that



[Mental states] are the subject matter of scientific laws, namely those comprising the science of psychology. Unfortunately, naturalism, so understood, is a doctrine not worthy of the name. The immediate problem is that it does not rule out the possibility that mental states are states of Cartesian souls. (Tye 1992, p.432)

Why should a philosopher genuinely concerned with science – which Tye and many others who share similar concerns claim to be – be vexed about such things? Cartesian souls, ESP, dowsing, sorcery, mesmerism, and so forth won't be admitted into our scientific ontology for no further reason than that they are useless to scientists in making progress on the problems driving their research; cf. Stich (1996, p.199). Questions of metaphysical relations to basic entities don't enter the picture. Given their usual characterization (e.g., the phenomenon in question balks at experimental control, it refuses to reveal itself to doubters), they are immediately disqualified from appearing in successful (promising, etc.) scientific theories.

The preceding observation is consistent with the claim that (1) and (2) each are necessary conditions on the naturalness of a kind, as a violation of (2) alone would be enough to return a verdict that a kind is not natural. However, there is good reason to suppose that if scientists somehow found that they could better control and predict the phenomena they study by including “weird stuff” in their models, they would do so, despite any uneasiness some might have about its physicalistic status. The legitimacy of such incorporations would ultimately be decided by the long-run successes or failures of research conducted around them, and not by (anti-) naturalizing projects as typically conceived by philosophers.

For any putative kind X, X either appears in our most promising theories and explanations or it does not. If the latter, then there is no basis for including X in our

scientific ontology. That is a happily naturalistic result. Suppose that X does figure in our best theories and explanations, but it chronically defies explanation by anything at a lower-level, raising concerns about “spooky action-at-a-distance”, “ectoplasm”, “mind stuff”, etc. The history of science indicates that the ontology of the sciences – and our rough-and-ready understanding of what counts as physical – will be extended to include X, regardless of how this offends our current sensibilities; cf. Chomsky (1968, pp.83-84; 2000, pp.144-145). This renders (1) an empty condition on scientific legitimacy. Eventually, if X continues to earn its theoretical keep and survives the stress of the rigorous tests imposed by actual scientific practice, what was once so reviled becomes an accepted part of the scientific worldview. This, too, looks to be a naturalistic outcome. At least with respect to scientific ontology, the justification for something’s being real typically comes from the work it does in our best scientific theories. If something plays a crucial role in our best scientific theories, we are ipso facto justified in treating it as real. Epistemology drives metaphysics: our best theories determine what we consider to be real. The idea that one has to advance a metaphysical thesis like physicalism in order to prevent the corruption of our scientific ontology is profoundly mistaken.<sup>6</sup>

A natural objection at this point might be, “Well, maybe that’s how scientists do things, but, if so, that’s an error on their part.” In other words, the story sketched so far might be fine as a description of scientific practice, but, normatively, it is dubious or impotent. These issues are admittedly quite complex and controversial, and are deserving of sustained discussion on their own. However, the groundwork for a compelling case for

---

<sup>6</sup> For a similar point about the metaphysics of color, cf. Johnson & Wright (2006).

giving the approach described above normative bite can be developed here. The next section's discussion of explanation further works out some related points.

A good place to begin is Ronald Giere's characterization of naturalism as involving a commitment to the idea that "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" (Giere 1999, p.5). Building out further, a mere description of actual scientific practice will not be blindly accepted as providing normative guidance. Rather, a normative account will best develop by examining the details of actual scientific practice and subjecting the methods (etc) discovered there to continual scrutiny, with additions, revisions, and expulsions carried out as needed. The key question has to do with the basis of that scrutiny. The appropriate standards of evaluation will include only demands that are recognizable as being conducive to furthering inquiry into the workings of the world around us; certainly, a concern for the advancement of our knowledge of the world ought to be at the core of the naturalistic enterprise. The naturalist will add that, as it happens, those demands are best shaped from the perspective of what our currently best scientific picture of the world licenses. Given that our currently best scientific picture of the world is incomplete and contains errors, the result is that we have to accept that any given moment, our best methods are fallible and in need of improvement; cf. Maddy (2007, esp. chapter 1). Importantly, the naturalist's endeavors to develop normative principles through direct engagement with our currently best science have the potential to aid empirical work, any successes of which will become available to the naturalist to further improve and refine her normative principles. One way of putting the central thesis of this paper is that naturalizing projects are not really naturalistic in the sense just outlined.

This approach is not intended to, and should not be expected to, deliver an account of unshakeable normative principles. Some might object to having to take on such uncertainty and perpetual revisability, and the evident circularity will also draw fire. A complaint likely to be lodged is that by some other set of standards developed (in some measure) apart from the sort of considerations called on above, this approach and its results look horrendous. Continuing this line, the undesirable consequences of this brand of mad-dog naturalism are precisely what should convince us that more traditional philosophical projects have a significant role to play. While it may be true that there are circumstances we would rather be in, it does not follow that any of them are available to us. It also doesn't follow that some altogether different method, one developed apart from our best current scientific theories, is capable of transporting us to one of those more desirable locales. To think that those things do follow is to accept an unjustifiedly optimistic (more likely, false) dichotomy.

To support her method for assessing and altering our normative principles, despite its admitted limitations, this radical naturalist can call upon the impressive record of success and progress delivered by scientific investigation. This is not to naively suggest that scientists have always faithfully or consciously employed sound methods, nor to conveniently ignore the scientific calamities that dot the historical record. In fact, an important task for naturalists of this stripe is to look at current and historical scientific practice to determine when and how scientists meet or fall short of standards, and to assess the respective benefits and penalties. Meanwhile, there has been a conspicuous lack of success and progress achieved by extrascientific means; again, the standard of evaluation in play has to do with what promises to advance inquiry into the workings of the world. Thus the burden of proof is on those who would stump for a means of generating a normative account that is in some way independent of actual scientific practice. Worth noting is that on

the account of naturalism sketched here, were some more traditional (i.e., armchair) philosophical method to prove itself as a facilitator of productive research, it would quickly get co-opted into the naturalist's suite of methods. These points can now be connected to the initial discussion.

Consider again circumstances in which (1) and (2) come apart. First, suppose that scientists continually failed, despite Herculean effort, to advance their research by making use of some posited entity X, but that a physicalistic account of X were available. Based on (2), we should not be naturalists about X; at best, X is on a par with glurg. It would be cold comfort to scientists who are prepared to jettison X to learn that we can be good physicalists about X. On the other hand, consider a case in which scientists, having analyzed their data using various techniques for extracting hidden structure, uncover the latent variables underlying some subset of the data. The hypothesis is formed that what the latent variables represent is X. Subsequently, scientists go on to include X in their models and they end up making considerable progress on the basis of their use of X; e.g., they have greatly increased their ability to predict and control several, perhaps related, phenomena by incorporating X in their models. However, as this is going on, philosophers conclude or strongly suspect that a physicalistic account of X is impossible. Would philosophers actually approach scientists and tell them that there is something illegitimate about developing scientific theories around X? If they did, should the scientists be expected to take such claims seriously? The earlier quoted remark from Fodor about spending the taxpayers' money gives a good indication that we can safely say that the answer to the first question is "Yes".

As for the second question, it is clear that a number of scientists take seriously philosophers' concerns about the scientific status of certain kinds (something often lamented by Daniel Dennett). However, the following comments from Francis Crick and Christof

Koch regarding the recent burst of philosophical theorizing about consciousness are likely an accurate representation of the sort of reaction one should expect from most scientists:

[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. For this reason neuroscientists should listen to the questions philosophers raise but should not be intimidated by their 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)

To be clear, Crick and Koch are expressing a dim view of the use of traditional philosophical methods to decide questions about the nature of worldly phenomena. It is fair to say that scientists will engage in metaphysical speculation of a sort when new hypotheses need to be generated and the evidence at hand doesn't suggest a clear-cut path for future research; it is vital, though, for that speculation to be recognized as such. For this reason, Crick and Koch are right to recommend that philosophers' questions are not to be dismissed out of hand. What they object to is resolving matters of concern to the scientist – such as rejecting or accepting hypotheses, or settling issues of ontological legitimacy – by a priori means. Their reason is that the results of past attempts to do so give us every reason to think that such a method is not up to the task. This calls into question the importance of the debate over the physicalistic status of X to scientific research.

Some have pointed out that arguments grounded in conceptual reflection run the risk of exposing little more than our current ignorance (Churchland 1997) and our “very basic inclination to overestimate our human capacities for anticipating the unexplored” (Wilson 2006, p.88). The proper significance of such observations is not that difficult problems are

sure to eventually yield to scientists' efforts and that our current ignorance is just a temporary stop on the sure path to knowledge. It is instead that scientific progress is extremely difficult to achieve and typically accumulates one little bit at a time, as part of a process involving trial and error, and not infrequently in the absence of any clear understanding of where inquiry will lead in the end. Thus the approach advocated here has as a virtue that it discourages a view of science as an inevitable, uninterrupted march toward certain knowledge. Drawing on the statistician John Tukey's remarks about the hazards of overreliance on confirmatory data analysis, forcing an a priori model onto inquiry – in the case of the naturalization debates, either verdicts about the nature of empirical phenomena reached through armchair reflection or normative models of scientific research formed independently of engagement with actual scientific practice – discourages exploration and hinders our chances of increasing our knowledge. It can be difficult to even figure out what are the right questions to ask about a particular phenomenon and what should be measured in its study, and certainty – which is required to slam the door on conceivability arguments – turns out to be an illusion throughout the sciences; cf. Tukey (1969, p.85). For these reasons, in a tug-of-war between the prospect of continued scientific progress on the basis of the inclusion of X in one's theories and yielding to worries about the physicalistic standing of X generated by what are largely a priori means, the latter is sorely out-muscled.<sup>7</sup>

---

<sup>7</sup> An insightful example from applied mathematics is Oliver Heaviside's operational calculus, which is discussed extensively in Wilson (2006). Heaviside treated differential operators as though they were numbers. That is precisely the sort of thing one is warned against very early in a standard calculus course. Heaviside's odd method was roundly criticized by his peers and the Royal Society refused to publish some of his findings. Nonetheless, he "invariably obtained correct answers [to applied problems] through such

A message one could extract from the passage from Crick & Koch is that it would be irresponsible or unreasonable for scientists to eliminate X from their promising theories on the basis of such worries. Continuing points already raised, this is not to recommend a blind fealty to scientific results, nor to suggest that all our questions will be answered by scientific investigation; scientists themselves caution against that. Rather, it is a recognition of what is routinely granted when it comes to the study of non-mental phenomena and for which there is no good reason to deny its extension to the “study [of] humans ‘above the neck’” (Chomsky 2000, p.76): the predictive and explanatory successes of our best scientific theories, and the methods and tools employed by practicing scientists, are invaluable guides in answering questions about the reality and nature of the phenomena they treat.

The earlier noted requirement that a posited kind continue to earn its keep is quite substantive. As inquiry proceeds, there will be many opportunities for posited kinds to prove themselves unworthy and thus be eliminated from scientific theories. For example, a posited kind might turn out to fare poorly when used to make predictions about new data sets and researchers in other fields might be repeatedly thwarted in their attempts to utilize the posited kind when investigating related phenomena in their own fields. It is routine for the nature of a posited kind to undergo significant development as inquiry proceeds and attempts are made to extend and ground its application. However, if the kind in question (or

---

apparent lunacy. Moreover, his algorithm generally found the right answer more quickly than orthodox methods (when the latter could be made to work at all)” (ibid., p.520). The impressive body of successes achieved by Heaviside’s methods eventually led electrical engineers to widely adopt them, despite the mysterious nature of their operation. Only decades afterward were the mathematical underpinnings to Heaviside’s methods finally completed (ibid., pp.530-31).



some suitably revised version of it) were to find no application outside the data set that initially prompted its being posited, scientists would ultimately determine that the kind is too fragile to maintain and it would be discarded. Claiming that epistemology drives metaphysics when it comes to these issues is not to advocate an “anything goes” approach, especially once it is recognized that the history of science indicates that scientists will go out of their way to try to “break” kinds that are introduced in this way; this attitude is part of the much celebrated self-correcting nature of the sciences. In a circumstance in which our “dodgy” X continues to do good work for us despite being exposed to strenuous testing, there is every reason to think that the right conclusion to draw is that either X has some physical realization, no matter how obscure it might currently be, or X should be added to our ontology in some more basic way. Such a conclusion could be reached while granting, as is always the case, that any number of future discoveries could force us to revise our commitment to X. Admittedly, there is no telling in advance exactly what events would have to unfold or how long it might actually take for posited kinds to become widely embraced within a field, but there is no reason to think a precise recipe is needed in order to justify this way of sorting things out.

Interestingly, Chalmers’ (2002) treatment of “Type-F Monism” – a variant of panpsychism he considers in the wake of his claim that consciousness cannot be physically explained – exploits a line of thinking similar to that explored here; one example he uses is electromagnetism in the Nineteenth Century. Type-F Monism has it that (proto-) phenomenal properties are the categorical bases of the dispositionally-characterized basic physical entities, which puts consciousness at the most fundamental level of reality. However, Chalmers (1996, p.126) is reluctant to count consciousness as a physical property, should it turn out to be on a par with or even deeper than the fundamental entities of physics.

While he chalks this up largely as a terminological dispute (ibid., pp.128-129), he also remarks that if physicalism could survive such a radical change as countenancing fundamental mental properties and laws, then it is “a weak doctrine indeed.” The consequent of that conditional looks like a promising horse to back, considering the difficulties that have dogged attempts to specify a scientifically relevant version of physicalism and the previous expansions of the domain of the physical to accommodate entities and processes that were troublesome when first considered. All this aside, I disagree with Chalmers about the prospects for explaining consciousness and thus about there currently being a reason to think that consciousness will have to be added to our scientific ontology in a basic or brute way, mainly because I reject the form of explanation he employs. That is the topic of the next section.

### **3 Explanation<sup>8</sup>**

As was suggested in section 1, a catalyst for much of the metaphysical theorizing about consciousness is the “hard problem” of explaining consciousness. The difficulty in comprehending how consciousness could ever be explained might create an opening for sparing the scientific import of naturalization projects, even if the points of the previous section are conceded. A nagging concern felt by many is that not only do we currently lack anything like a satisfactory explanation of consciousness, no-one has the foggiest idea of how consciousness could be fully incorporated into the scientific worldview without radically altering our inventory of scientifically legitimate kinds; e.g., adding consciousness

---

<sup>8</sup> This section’s discussion of explanation is developed further in Wright (2007). It should be noted that Trout (2002) is a starting point for both this section and that paper.

to our ontology in some fundamental way. The thinking is that consciousness seems so peculiar, so unlike anything else that has yielded to scientific explanation, that there is little hope that we could ever be in a position to fully account for it in terms of the organization and activity of anything else. While acknowledging the past expansions of the domain of the physical, it is understandable to be distressed by the prospect of having to accept that the universe is so strange that consciousness is an elementary strand in the fabric of reality.

Would-be naturalizers can thus be read as doing nothing more than trying to build a case for the possibility of a genuine explanation of consciousness, so that we are not obliged to expand our scientific ontology in an extreme way; a robust commitment to the importance of physicalism need not be taken as central to those endeavors. To the extent that their metaphysical theorizing helps us see a deeper continuity between phenomenally conscious states and the kinds invoked in various branches of science – reducing the sense of dumbfoundedness that descends upon one when trying to comprehend just how it is that anything like neural activity or information processing could be entirely responsible for qualitative experience – they will have been successful in that effort. Would-be anti-naturalizers, then, are to be taken as arguing that that undertaking is hopeless and, as a consequence, we will never achieve a satisfactory explanation of consciousness. The entire debate over the hard problem, however, turns on a misconception of explanation. With a more appropriate form of explanation in place, it is clear that solving the hard problem is not a requirement for a successful explanation of consciousness and thus the driving force behind naturalization projects disappears.

The following does a good job of bringing out the notion of explanation at work in the hard problem:

For almost every natural phenomenon above the level of microscopic particles, there seems in principle to exist a reductive explanation.... when we give an appropriate account of lower-level 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)

Here is another helpful example:

[In] showing how an 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)

Consider also:

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)

One last example:

[Chalmers & Jackson (2001)] can claim, with some plausibility, that we will only ever be satisfied with a proposed reduction [as part of an explanation] when the micro-phenomena mesh in the right way with the concepts used to characterize the target, in such a way as to warrant an a priori conditional. It is only when we can see that changing concentrations of hydrogen ions will produce just the kinds of changes distinctive of a changing pH value, that we will accept that the latter is constituted by

the former. (Carruthers 2004, p.160; the first emphasis is mine, the second emphasis is in the original)

Similar quotes are easy to find in the consciousness literature. The thinking is that once we have the low-level physical details in place, the higher-level phenomenon convincingly appears to be a predictable consequence. The use of expressions such as “falling out”, “sense of mystery”, “meshing”, and “seeing why” strongly suggests that reflection on a good explanation should elicit a “sense of understanding” or a “feeling of knowing” (Trout 2002), a warm glow that signals that one really understands the target phenomenon and how it comes to be; i.e., what really follows from the explanatory factors.

One way of putting the worries about explaining consciousness is, “Look, tell me all you want about neural activity, microtubules, information processing, or anything else like that. None of that gives me the first clue about how it is that my pains feel the way they do, that sunsets look to me the way they do, or why I even have qualitative experience. Those aren’t even close to being the right kind of things that would get me to not be utterly mystified by consciousness.” Many who take seriously that the hard problem has consequences for the scientific study of consciousness often point out that one does not run into similar reactions when considering cases of successful scientific explanation. A popular example is the explanation of boiling afforded by the chemical theory of water. After sketching an account of the motion of H<sub>2</sub>O molecules, intermolecular forces, vapor pressure and so forth, Joseph Levine (1993/1997, p.549) concludes that it is “inconceivable that H<sub>2</sub>O should not boil at 212°F at sea level.” In the case of consciousness, we do not, and very likely cannot, have a satisfactory explanation because it is a simple matter to come up with ways to cast doubt on the connection between the explanans and explanandum.

The reliance on a sense of understanding in the model of explanation at play in discussions of the hard problem is highly suspect. The sense of understanding provides neither a necessary nor a sufficient condition on explanation. In uncritically accepting a model of explanation that includes an appeal to epistemic phenomenology, philosophers – and no small number of scientists – have not just set the bar too high. They have also overlooked or marginalized traits that are vitally important to what counts as genuine understanding and a good explanation.

One problem for the sense of understanding that eliminates it as a sufficient condition on explanation is that epistemic phenomenology fails abysmally as a guide to truth. Trout (2002) argues that the sense of understanding does not permit us to recognize good explanations in its presence. The history of science is brimming with instances in which an overwhelming sense of understanding accompanied a faulty explanation. As the appeals to “epistemic necessitation” and “falling out” make plain, the sense of understanding is intertwined with a conviction that the target phenomenon is an inevitable consequence of the explanatory facts. To use Carruthers’ expression, the explanans and explanandum mesh so well that it does not seem possible that things could be other than they are. In no shortage of cases, this conviction feeds the impression that one’s explanatory work is done, although the accepted explanation might actually be quite incomplete or flat wrong.

Ptolemy exemplifies the irrelevance of the sense of understanding. Much of the strength of his sense of understanding of his false theory is attributable to how well it cohered with the rest of his incomplete and error-riddled body of beliefs. Not only did it steer him to a false theory, the sense of understanding led Ptolemy to ignore alternatives to his geocentric explanations and to refrain from pursuing deeper investigation into core

principles of his theory; cf. Trout (2002, pp.223-224).<sup>9</sup> His sense of understanding was so great that in the face of difficulties for his explanation of the motions of heavenly bodies, Ptolemy added to it ever-more complicated details, rather than scrap it altogether. Errors like Ptolemy's have been repeated throughout the history of science. Even those who end up on the right side of a debate are not immune to these errors. Should the feeling of knowing accompany a good explanation, it is still possible for the strong sense of "fit" to lead to questionable practices, such as dismissing competitors to a good explanation for the wrong reasons; cf. Trout's (2002, p.227) criticisms of Copernicus' "shuddering" at the thought of alternatives to heliocentrism. The recurrence of such errors across many different individuals and circumstances indicates that one need not proceed with mendacity in order to vigorously defend a false theory with little supporting evidence or in the face of countervailing evidence, or to ignore a perhaps legitimate challenge to a thus-far successful theory. The structure of our minds is such that we can be led to overestimate our accomplishments and deprecate potential underminers of our beliefs, even when we have the best of intentions.

Trout (2002) cites two well-researched psychological biases – those of hindsight and overconfidence – as the source of the powerful intuition that the feeling of understanding

---

<sup>9</sup> It is important to note, especially in connection with points that will be discussed shortly, that the complaint about Ptolemy has to do with the mechanisms that are supposed to govern the motions of heavenly bodies, not his theory's predictive adequacy. Contemporary models of planetary motion share many important features with Ptolemy's; cf. Saari (2005, p.12).

somehow clues us into an explanation's worth.<sup>10</sup> Those biases make us prone to strongly hold false beliefs about how well we understand something and what outcomes must follow from cited explanatory factors; the latter is especially relevant to claims about "seeing the fit" and epistemological necessity. Also, once a strongly held false belief is in place, it might be terribly hard to shed due to other biases; e.g., the confirmation bias.<sup>11</sup> In the light of the impressive body of empirical evidence that calls into question our powers of prediction, evidence-handling, and gauging the extent of our knowledge, it is puzzling that objective measures of explanatory adequacy would be put on a par with, and sometimes even made subordinate to, subjective measures. Even if we were to have before us an explanation that allows us to "see" the relation between its explanans and explanandum, that alone would offer no assurance that we should accept the explanation. The sense of understanding sought in the hard problem may provide some intellectual consolation to us. As Daniel Schacter (2001, p.149) remarks about the hindsight bias, "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." When such good feelings lead us to misdirect or prematurely halt our inquiry, they harm our efforts to learn more about the world.

A common reaction to it being pointed out that the sense of understanding has a poor track record as a guide to truth, is to acknowledge that of course at times thinkers have been

---

<sup>10</sup> For the hindsight bias, cf. Dawes (1999), Fischhoff & Beyth (1975), Hoffrage et al (2000), Schacter (2001, pp.145-149). For the overconfidence bias, cf. Fischhoff et al (1977), Henrion & Fischhoff (1986), Lichtenstein et al (1982).

<sup>11</sup> Cf. Nickerson (1998). Rabin & Schrag (1999) offer an extended argument for the claim that the confirmation bias induces overconfidence. It is plausible that those two biases are mutually reinforcing.



led astray by the gripping epistemic phenomenology that accompanies reflection on what turns out to be a bad explanation. However, the response continues, it is implausible to suppose that the sense of understanding is not a necessary condition on satisfactory explanation. The idea is that we might not be able to tell that we have a good explanation on the basis of the presence of the warm glow of understanding, but we can be certain that we have a bad explanation on our hands if that feeling does not appear. This is mistaken.

A crucial point that has been overlooked in the debate over the hard problem is that an account of what we should look for in a good explanation, at least when it comes to many areas of psychology and neuroscience, will prominently feature control of the target phenomenon. As Colin Camerer (2003, p.160) puts it, “Psychologists feel that they understand a phenomenon when they can ‘turn it on and off’.” Douglas Tweed (2003, p.14) has commented that a great deal of neuroscientific inquiry is dedicated to achieving “control. We want the power, not just to turn out new minds, but to specify the details, so that, ultimately, we can repair and improve the minds we have.” Carl Craver (forthcoming) has called attention to the centrality of control in the neurosciences and other fields, and issues of manipulation and intervention figure prominently in James Woodward’s (2003) work.

Consider the following sketch of an account of explanation that makes control a chief concern: 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 were things different. Explanations along these lines go a long way toward uncovering the structure of target phenomena in a manner that advances scientific inquiry. A great deal of work will go into, for example,

tracking down a mechanism, decomposing it into its parts and their various relations and activities, and devising techniques 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 the much sought-after sense of understanding might often accompany such explanations (e.g., "Oh, I see how that works"), there is no reason to think that it must do so or that it can be used to adjudicate between true and mistaken claims about manipulability.

Explanations that make so much of manipulation might be unsatisfying in that there is no assurance that we will always be able to "see" the relation between a mechanism and the target phenomenon. They are, however, 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 a feeling of understanding that would meet the challenge of the hard problem, there is an authentic, deep form of understanding that comes with a good control explanation. These explanations go well beyond prediction alone and mere simulation, 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, for example, predicting and retrodicting the behavior of the phenomenon. That achievement has considerable explanatory force and it has an epistemic payoff far greater than, for example, a systematized description of experimental results or a laundry list of correlations. It is unclear what further "mystery removal" scientists should be tasked with.

In connection with the preceding section's discussion, having a good control explanation in hand equips scientists 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 advance their investigation of related phenomena. Considering all the virtues that explanations can have without it and the difficulties reliance

on it can engender, epistemic phenomenology is poorly suited as a requirement on scientific explanation. There is perhaps a natural expectation that we should feel like we understand things that we have explained. In cases in which the sense of understanding does not dawn on us upon first encountering a proposed explanation, if the explanation holds its own over a long period of time and becomes a familiar part of the scientific canon, that might eventually lead to the sense of understanding being induced in those who reflect on it; the details of whether this is so and how it might come about are topics to be explored in the cognitive science of science.<sup>12</sup> That the sense of understanding might eventually attach to an explanation that becomes well-entrenched, though, is irrelevant to how we should evaluate any explanation that comes before us.

Ultimately, the sense of understanding provides neither a necessary nor a sufficient condition on satisfactory explanation. This takes the sharpest teeth out of the problem of explaining consciousness, but there is no guarantee that consciousness will be explained with a better form of explanation in place. In dispelling the hard problem and emphasizing the role of control in explanation, I do not mean to imply that it will be easy to identify control mechanisms for consciousness or that such mechanisms certainly exist. Rather, I am claiming that the debates between philosophers about the hard problem have no bearing on the prospects for a scientific explanation of consciousness. Of course, there might be other, perhaps insurmountable, obstacles to explaining consciousness that have nothing to do with the hard problem; e.g., ethical considerations or technological barriers that limit

---

<sup>12</sup> For more here, cf. Trout (2007). How such research plays out may also suggest further replies to anti-physicalists who play up the vast apparent difference between past explanatory successes and any explanation of consciousness that might be offered in the future.

experimentation. Jaegwon Kim (1996, p.177) makes a distinction between what is needed to achieve a scientific understanding of consciousness and what is needed to solve the philosophical problem of consciousness. My suggestion is that Kim is closer to having the right view of the situation than are Chalmers, Jackson, Levine, and the many others who write as though the form of explanation at issue in the hard problem is the kind we should expect our best scientific work to deliver. The key point is that the science of consciousness should not be counted on to – and should not be considered in any way deficient if it fails to – meet the challenge of the hard problem that has garnered so much attention from philosophers. And with that, a major motivation for thinking that the metaphysics of consciousness is scientifically relevant is also banished.

Two possible replies merit discussion. First, one might challenge the portrayal of the hard problem as appealing to epistemic phenomenology. To the simplest way of putting such an objection, the best response is to point to the earlier cited claims from leading figures in the consciousness literature. A more sophisticated version of the objection, however, takes the appeals to epistemic phenomenology to be mainly rhetorical in nature. Talk of epistemic phenomenology is useful for getting people to recognize that there is a hard problem of explaining consciousness, but the hard problem itself is generated by a host of arguments that turn on intuitions about the possibility of zombies, inverted spectra, and so forth; epistemic phenomenology isn't really central to the hard problem.

In reply to this more sophisticated variant, it is fair to ask about the admissibility of the intuitions that are generating the “real” (i.e., not epistemic phenomenology-dependent) hard problem when evaluating a candidate explanation. Why should they be granted any weight at all? A non-starter would be to say that the intuitions are very compelling and that there is good reason – perhaps based on reflection on past experience – to suppose that

anything that seems so gripping should be taken seriously as indicative of the facts of the matter or live possibilities. This is just to smuggle back in the discredited appeal to epistemic phenomenology that it is the express purpose of the maneuver under consideration to avoid. One might sidestep an illicit appeal to epistemic phenomenology by insisting that it is the machinery of “cold” conceptual reflection alone that is to do the needed work. The intuitions follow directly from what is and is not contained in our concepts and from the ways in which certain concepts are and are not connected to one another.

The thoroughgoing naturalist introduced before is going to be very interested in this method’s prospects for contributing to inquiry. The answer, echoing remarks from section 2, is that it does not have a particularly good track record at settling questions about the workings of the world. While one might still opt to employ armchair methods, despite their poor showing at delivering insight into the nature of empirical phenomena, such a policy threatens to drive a deep wedge between philosophical and scientific activity. Given its past performance, the armchair philosopher surely has the burden of arguing for the bearing of her extrascientific method on the assessment of scientific activity. Relevant here are Stephen Stich and Jonathan Weinberg’s (2001) concerns about the use of armchair reflection to “define the subject matter” in order to judge the adequacy of scientific theories, to ensure that the right phenomenon has been explained, etc; cf. Jackson (1998). Based on a review of relevant empirical research (including data from their own experiments), they argue that such a project cannot maintain its assumptions about the nature of concepts and the typicality of one’s own intuitions. Taking into account both the well-researched limitations and peculiarities of human psychology and the “track record argument” against armchair methods, there is a solid basis for concluding that genuinely promising ways of evaluating and furthering scientific inquiry will be firmly rooted in the results and methods of actual

scientific practice. Unlike accounts of explanation that incorporate a priori methods or epistemic phenomenology, the account sketched before, with its emphasis on control, is already well-positioned to contribute to advancing research.

The second objection is that the discussion has naively idealized actual scientific practice, as scientists are made out to be far more objective than philosophers. A casual review of the history of science makes plain that scientists can fail to meet standards of objectivity when (e.g.) deciding which hypotheses to accept or reject. This objection would be damaging to a naturalistic approach that blindly relied on reports of actual scientific practice. As was made evident in the last section, however, the strain of naturalism embraced here involves a continual examination of one's methods in the light of new developments. Our naturalist is not concerned merely with how things are done, but with how things ought to be done, given the best current picture of the world. Scientists and philosophers are equally subject to criticism when their practices do not meet the standards uncovered by the naturalist in her diligent methodological churnings.<sup>13</sup> That scientists are sometimes influenced by things that philosophers are criticized for being influenced by nicely illustrates the need for the sort of naturalistic approach being championed. Furthermore, the workaday concerns of many scientists typically center on a small cluster of very specific problems. This creates an obvious role for naturalist philosophers to play in cultivating and enforcing methodological hygiene, by, inter alia, looking out across a broader swath of research and thinking about problems not directly tied to the pressures

---

<sup>13</sup> Two relevant examples of things that might bear on our naturalist's activities are that some of the psychological research on biases cited before was conducted on scientists making claims about their own areas of expertise and that there is no shortage of critical discussion by measurement theorists and statisticians of scientists' statistical practices.

attached to the next run of experiments. To dredge up another point from section 2, it would be irresponsible or unreasonable for a scientist to reject the naturalist's normative advice, were it grounded in solid evidence gathered from our best ways of finding out about the world and its potential benefits to the scientist's research – relative to those of the scientist's current methods and other methods she may have to choose from – clearly demonstrated.

#### **4 Conclusion**

The preceding discussion might strike some as arguing that philosophical discussions of consciousness are entirely without value. My purpose, however, has only been to demonstrate that there is a greater need to carefully delimit philosophical work on consciousness that is justified by its expected contribution to scientific research and that in need of extrascientific justification. Thus this paper is an invitation for those involved in the debates over the physicalistic status of consciousness to fully spell out the expected payoff of their work. That might involve motivating the significance of extrascientific concerns about consciousness or, of course, responding to the challenges herein regarding the scientific relevance of the standard debates over physicalism and the hard problem. The paper is also an invitation to philosophers interested in consciousness to consider another way of trying to make a contribution to the sciences, if it turns out that they cannot defend the scientific relevance of their current projects. These undertakings are important because, although the issues addressed in the consciousness literature admittedly seem deeply interesting and puzzling (e.g., Is physical stuff really the basis of something that feels like this?), other issues in the past have struck philosophers the same way but have turned out to be based on bad questions; e.g., How many angels can fit on the head of a pin?, Does adding

three to five increase the three or the five to eight? This is not to say that the core philosophical issues of consciousness are pseudoproblems, but to raise the concern that it is not clear why they are not pseudoproblems.

## 5 References

- Camerer, C. (2003). Behavioral Game Theory. Princeton: Princeton University Press.
- Carruthers, P. (2004). "Reductive explanation and the explanatory gap." Canadian Journal of Philosophy, 34, 153-174.
- 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.
- \_\_\_\_\_. (2004). "How Can We Construct a Science of Consciousness?" In M. Gazzaniga (ed.) The Cognitive Neurosciences III. Cambridge, MA: MIT Press.
- Chalmers, D. & F. Jackson. (2001). "Conceptual Analysis and Reductive Explanation." Philosophical Review, 110, 315-360.
- Chomsky, N. (1968). Language and Mind. New York: Harcourt Brace Jovanovich.
- \_\_\_\_\_. (2000). New Horizons in the Study of Language and Mind. Cambridge: Cambridge University Press.
- Churchland, P.S. (1993/1997). "Can Neurobiology Teach Us Anything about Consciousness?" In N. Block, O. Flanagan, G. Guzeldere (eds.) The Nature of Consciousness: Philosophical Debates. Cambridge, MA: MIT Press.
- \_\_\_\_\_. (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.
- Crook, S. & C. Gillett. (2001). "Why Physics Alone Cannot Define the 'Physical': Materialism, Metaphysics, and the Formulation of Physicalism." Canadian Journal of Philosophy, 31, 333-360.



- 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.
- 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.
- Fodor, J. (1975). The Language of Thought. Cambridge, MA: Harvard University Press.
- \_\_\_\_\_. (1987). Psychosemantics. Cambridge, MA: MIT Press.
- \_\_\_\_\_. (1990). "Roundtable Discussion." In P. Hanson (ed.) Information, Language, and Cognition. Vancouver: University of British Columbia Press.
- Giere, R. (1999). Science Without Laws. Chicago: University of Chicago Press.
- Gillett, C. (2001). "The Methodological Role of Physicalism: A Minimal Skepticism." In C. Gillett & B. Loewer (eds.) Physicalism and Its Discontents. Cambridge: Cambridge University Press.
- Greene, B. 1999. The Elegant Universe. New York: W.W. Norton & Co.
- Hempel, C. (1969). "Reduction: Ontological and Linguistic Facets." In S. Morgenbesser, et al. (eds.) Philosophy, Science, and Method. New York: St. Martin's.
- 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.
- Holman, E. 2006. "Dualism and secondary quality eliminativism: Putting a new spin on the knowledge argument." Philosophical Studies, 128, 229-256.
- Jackson, F. (1982). "Epiphenomenal Qualia." Philosophical Quarterly, 32, 127-136.
- \_\_\_\_\_. (1986). "What Mary Didn't Know." Journal of Philosophy, 83, 291-295.
- \_\_\_\_\_. (1994). "Armchair metaphysics." In J. O'Leary-Hawthorne & M. Michael (eds.) Philosophy in Mind. Dordrecht: Kluwer.
- \_\_\_\_\_. (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.
- Kim, J. (1996). Philosophy of Mind. Boulder, CO: Westview Press.
- 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.
- Lichtenstein, 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.
- Maddy, P. (2007). Second Philosophy. Oxford: Oxford University Press.
- Melnyk, A. (1997). "How to Keep the 'Physical' in Physicalism." Journal of Philosophy, 94, 622-637.
- Nickerson, R. (1998). "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." Review of General Psychology, 2, 175-220.
- Poland, J. (1994). Physicalism. Oxford: Clarendon Press.
- Rabin, M. & J. Schrag. (1999). "First Impressions Matter: A Model of Confirmatory Bias." Quarterly Journal of Economics, 114, 37-82.
- Robinson, W. (2004). Understanding Phenomenal Consciousness. Cambridge: Cambridge University Press.
- Saari, D. (2005). Collisions, Rings, and Other Newtonian N-Body Problems. Providence (RI): American Mathematical Society.
- Schacter, D. (2001). The Seven Sins of Memory. Boston: Houghton Mifflin.
- Schaffer, J. (2003). "Is there a fundamental level?" Nous, 37, 498-517.
- Stich, S. (1992). "What is a Theory of Mental Representation?" Mind, 101, 243-261.
- \_\_\_\_\_. (1996). "Naturalism, Positivism, and Pluralism." In S. Stich Deconstructing the Mind. New York: Oxford University Press.
- Stich, S. & S. Laurence. (1996). "Intentionality and Naturalism." In S. Stich Deconstructing the Mind. New York: Oxford University Press.

- Stich, S. & J. Weinberg. (2001). "Jackson's Empirical Assumptions." Philosophy and Phenomenological Research, 62 637-644.
- Suppes, P. (1978). "The Plurality of Science." PSA 1978, 2, 3-16.
- Teller, P. (2001). "Twilight of the Perfect Model Model." Erkenntnis, 55, 393-415.
- Trout, J.D. (2002). "Scientific explanation and the sense of understanding." Philosophy of Science, 69, 212-233.
- \_\_\_\_\_. (2007). "The Psychology of Scientific Explanation." Philosophy Compass, 2/3, 564-591.
- Tukey, J. (1969). "Analyzing data: Sanctification or detective work?" American Psychologist, 24, 83 - 91.
- Tweed, D. (2003). Microcosms of the brain: What sensorimotor systems reveal about the mind. New York: Oxford University Press.
- Tye, M. (1992). "Naturalism and the Mental." Mind, 101, 421-441.
- \_\_\_\_\_. (1994/1997). "A Representational Theory of Pains and Their Phenomenal Character." In N. Block, O. Flanagan, G. Guzeldere (eds.) The Nature of Consciousness: Philosophical Debates. Cambridge, MA: MIT Press.
- \_\_\_\_\_. (1995). Ten Problems of Consciousness. Cambridge, MA: MIT Press.
- \_\_\_\_\_. (2000). Consciousness, Color, and Content. Cambridge, MA: MIT Press.
- van Fraassen, B. (1996). "Science, Materialism, and the False Consciousness." In J. Kvanvig (ed.) Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge. Lanham, MD: Rowman and Littlefield.
- \_\_\_\_\_. (2002). The Empirical Stance. New Haven: Yale University Press.
- Wilson, M. (2006). Wandering Significance: An Essay on Conceptual Behavior. Oxford: Clarendon Press.
- Woodward, J. (2003). Making Things Happen: A Theory of Causal Explanation. Oxford: Oxford University Press.
- Wright, W. (2007). "Explanation and the hard problem." Philosophical Studies, 132, 301-330.