THE JOURNAL OF PHILOSOPHY

VOLUME CVII, NUMBER 5
MAY 2010

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

All communications to the Editors and Trustees and all manuscripts may be sent to Alyssa Timin, Managing Editor, at submissions@journalofphilosophy.org.

You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2010

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Libraries and Institutions (print and online)	$180.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc., MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027. Periodicals postage paid at New York, NY, and other mailing offices.

POSTMASTER: Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

ISSN 0022–362X

THE JOURNAL OF PHILOSOPHY

VOLUME CVII, NO. 5, MAY 2010

LESSONS FROM THE CONTEXT SENSITIVITY OF CAUSAL TALK*

Suppose we have a theory of singular causation according to which

(1) Caesar's birth was a cause of his death.

is true.[1] Charge: It offends common sense to say that Caesar's birth was a cause of his death. Response: The assertibility conditions of causal claims are affected by conversational context. Even if (1) is true, in normal contexts it will be uninformative, or misleading, or not a suitable answer to the sorts of questions we are interested in. And general pragmatic principles explain why it would offend common sense to assert even true sentences that are uninformative, misleading, or not topical. So it is no mark against a theory of causation that it predicts that (1) and certain other odd-sounding sentences are true.

This response, prominent in the work of Lewis,[2] Bennett,[3] and others, is based on the plausible idea that some distinctions made in natural language need not—indeed, should not—be reflected in metaphysics. Natural language does distinguish between Caesar's birth and Brutus's stabbing, with respect to being a cause of Caesar's death, but perhaps our metaphysics of causation should not. If we pursue this line, as I think we should, then we must ask which natural

* For helpful discussion and comments on earlier drafts, thanks to Ned Hall, Sarah Moss, Bob Stalnaker, Judy Thomson, Steve Yablo, and an anonymous referee for this JOURNAL.

[1] For examples of such theories, see J. L. Mackie, *The Cement of the Universe: A Study of Causation*, 2d ed. (New York: Oxford, 1980); David Lewis, "Causation," this JOURNAL, LXX, 17 (October 1973): 556–67; Lewis, "Causation as Influence," this JOURNAL, XCVII, 4 (April 2000): 182–97; and Jonathan F. Bennett, *Events and Their Names* (Indianapolis: Hackett, 1988).

[2] Lewis, "Causation," pp. 558–59; and "Causation as Influence," p. 196.

[3] Bennett, *The Act Itself* (New York: Oxford, 1995), see pp. 130–33.

0022-362X/10/0705/221–242

language distinctions do constrain our metaphysics, and how. These questions are especially important for distinctions that are sensitive to features of conversational context, because we should not inadvertently impute the effects of such context sensitivity to our metaphysics.

This paper starts by arguing that ordinary causal talk is far more sensitive to conversational context than has been recognized to date. I then formulate a principle that helps characterize that context sensitivity. I argue that this principle explains why some putative overgenerated causes are never felicitously counted, in conversation, as causes, and I argue that a plausibly strengthened version of the principle explains at least some of the oddness of 'systematic causal overdetermination'. These explanations are a natural extension of the line that Lewis, Bennett, and others take with (1): when we are confronted with linguistic data that threaten to make trouble for our metaphysics, we try to give a plausible explanation of the data that does not require any changes to our metaphysics. When we are successful, we are not obligated to change the metaphysics. The linguistic explanations of (1) and many other examples are often seen as a boon to metaphysics—a ground-clearing prerequisite to serious theorizing about causation. But the kind of context sensitivity I consider here has a worrying flip side. The metaphysics of causation turns out to be much less directly constrained by ordinary language judgments than we might have expected.

I. THE CONTEXT SENSITIVITY OF CAUSAL TALK

It is often noted that which events count as the causes of another event, in a conversation, is a dramatically context-sensitive matter. In light of this, many philosophers suggest that our intuitions about causation should not be influenced by judgments about sentences of the form '*c* was among the causes of *e*'. We should instead restrict our attention to putatively less context-sensitive sentences of the form '*c* was a cause of *e*'. This suggestion is underwritten by a tacit argument by analogy. Which books count as the books, in a conversation, is a dramatically context-sensitive matter. But even if what counts as a book is a little context sensitive, 'a book' is much less context sensitive than 'the books'. By analogy, 'a cause of *e*' is much less context sensitive than 'the causes of *e*'. Indeed, it is sufficiently insensitive to context to be a suitable guide when we work on the metaphysics of causation. This argument is seductive, but it is specious, at least because the analogy between 'a book' and 'a cause of *e*' fails.[4]

[4] Lewis does concede that "even 'a cause of' may carry some hint of selectivity" ("Causal Explanation," in *Philosophical Papers*, vol. 2 (New York: Oxford, 1986),

The analogy fails because a given definite description of the form 'the *F*s of DEFINITE NP' is generally much more specific than its 'the *F*s' counterpart: 'the friends of Liem', for example, is much more specific than 'the friends'. This specificity tempers the impact that facts about contextual salience can have on the interpretation of 'the *F*s of DEFINITE NP'. So definite descriptions whose nominals have genitive modifiers are generally less context sensitive than their counterparts without such modifiers—unless there is another reason for their context sensitivity.[5] (Continuing the example, 'the friends of Liem' is less context sensitive than 'the friends'.) This suggests that the dramatic context sensitivity of 'the causes of *e*' may not be fully accounted for by the fact that it is a definite description. And if its context sensitivity has another source, then it would not be surprising to find that 'a cause of *e*' is context sensitive, too.

In fact this is what we do find. Unfortunately, moreover, neglect of the context sensitivity of 'a cause of *e*' has led our theorizing about causation astray. To bring this out I want to look at some linguistic data that clearly should not be accounted for in our metaphysics. For this reason, these data put nonnegotiable demands on the linguistic theory that interfaces between our metaphysics and our ordinary language judgments: the linguistic theory has to account for these data on its own. But as we will see, it is plausible that a theory powerful enough to do this work can also do work usually taken to be the metaphysician's responsibility.

For all I will say here, the kind of context sensitivity that matters for present purposes may or may not make a difference to the truth conditions of causal claims. What I am interested in for present purposes

---

pp. 214–40, see p. 216). Peter Unger argues that the verb 'cause' and "other transitive causal verbs" are context sensitive in his "The Uniqueness in Causation," *American Philosophical Quarterly*, XIV, 3 (July 1977): 177–88 and *Philosophical Relativity* (Minneapolis: Minnesota UP, 1984). But Unger does not discuss 'a cause of *e*'. See Judea Pearl, *Causality: Models, Reasoning, and Inference* (New York: Cambridge, 2000), pp. 222–23 for contrast.

[5] It is interesting that most of Bertrand Russell's examples of definite descriptions have nominals with genitive modifiers. He singles them out as "descriptive functions," including "the father of *x*," "the sine of *x*," "the present King of France," "the author of Waverley," "the centre of mass of the Solar System at the first instant of the twentieth century," and so on (Russell, "On Denoting," in Gary Ostertag, ed., *Definite Descriptions: A Reader* (Cambridge: MIT, 1998), pp. 35–49, see p. 35 [Originally published in *Mind*, XIV, 56 (October 1905): 479–93]; and "Descriptions," in Peter Ludlow, ed., *Readings in the Philosophy of Language* (Cambridge: MIT, 1997), pp. 323–34, see p. 323 [Originally published in *Introduction to Mathematical Philosophy* (New York: MacMillan, 1919), pp. 167–80].) Definite descriptions of this form seem to encourage 'attributive' over 'referential' readings, in something like Donnellan's senses (Cf. Keith Donnellan, "Reference and Definite Descriptions," *Philosophical Review*, LXXV, 3 (July 1966): 281–304).

is how the assertibility conditions of sentences of the form '*c* was a cause of *e*' are sensitive to conversational context. By the *assertibility conditions* of a sentence, I mean the circumstances in which it is appropriate for a speaker who knows all the relevant nonsemantic facts to use that sentence. There is no doubt that the assertibility conditions of any given causal claim vary with conversational context, in the familiar way that the assertibility conditions of any sentence vary with conversational context: clearly it is often inappropriate to say rude things, or things that have already been said, or things that are manifestly obvious, and so on. What I want to call attention to here is one unnoticed way in which the assertibility conditions of '*c* was a cause of *e*' vary with conversational context.

Suppose I know that the leak caused the puddle, and that the puddle together with the cold caused the ice. I tell you about the ice, and you ask about its causes. I could begin to answer your question with any of:

(2) The leak was a cause of the ice. (...So if we fixed the leak...)
(3) The puddle was a cause of the ice. (...So if we fixed the drain...)
(4) The cold was a cause of the ice. (...So if we fixed the heater...)
(5) The leak was a cause of the ice, and the cold was a cause of the ice.
(6) The puddle was a cause of the ice, and the cold was a cause of the ice.

But I could not appropriately say

(7) # The leak was a cause of the ice, and the puddle was a cause of the ice.
(8) # The puddle was a cause of the ice, and the leak was a cause of the ice.

Assertions of (7) and (8) would not appropriately describe the stipulated causal relations. Indeed, they would misdescribe them, conveying that while the leak and the puddle were both causes of the ice, the leak was not a cause of the puddle.

What is wrong with using (7) and (8) to describe the stipulated situation? Roughly speaking, (7) and (8) wrongly double-count a causal route to the ice. Two bits of terminology will make it easier to give a more precise and more general explanation than this one. First, many philosophers of causation, including Lewis and Bennett, explicitly aim to characterize a "broad and nondiscriminatory" causal relation.[6] Such accounts say, for example, that throwing a switch so that a train continues down the right-hand track (instead of the left-hand track) bears this nondiscriminatory causal relation to the

[6] Lewis, "Causation," see p. 559.

train's arrival, even if the tracks rejoin before the arrival and the switching makes no difference to the time and manner of the arrival. There is significant disagreement about the nature of this relation, but I will assume that there is a metaphysical natural kind that Lewis, Bennett, and similar philosophers of causation are aiming to characterize. For convenience I will call this putative natural kind *causal relevance*. I leave open the question which analysis of causal relevance is right, and I also leave open the question which causal relata that are causally relevant to e can count as a cause of e in a given context. (Note that causal relevance is a directed relation, because the switch is causally relevant to the arrival but not vice versa.) Second, a (possibly infinite) sequence of causal relata $\langle \ldots, e_{n-2}, e_{n-1}, e_n \rangle$ constitutes a *causal path* to e_n just in case e_{n-1} is causally relevant to e_n, e_{n-2} is causally relevant to e_{n-1}, and so on. It will be important to confine our attention to causal paths that are *maximal* in the sense that they have no proper supersequences with the same terminal event. The "broad and nondiscriminatory" nature of causal relevance means that there will be very many causal paths to any effect, and that very many of those paths will overlap. It is because of this overlapping that I use the term 'causal path' instead of the more familiar 'causal chain': on my definition, distinct causal paths to e can overlap by sharing events other than e. Some intuitive understandings of the chain metaphor obscure this possibility.

I will appeal to the following principle in explaining why (7) and (8) misdescribe the situation we have been considering:

USE GOOD REPRESENTATIVES

When you ascribe some causal responsibility for e to a causal path to e, use good representatives of that path for the purposes at hand.[7]

The refinement of this principle would be an illuminating project, but fully characterizing the kind of context sensitivity it approximates would require intensive empirical investigation. My intention here is to sharpen the principle just enough to get at some ways in which our theorizing about causation should be informed by the context sensitivity of causal talk.

Before I explain how USE GOOD REPRESENTATIVES sheds light on the infelicity of (7) and (8), I need to say a bit about the thought that

[7] A serious discussion of absences and omissions would take us very far afield, but I want to note that it is easy to generate analogues of the leak/puddle case for absences. (This is a 'new' kind of context sensitivity, not obviously related to the broadly normative context sensitivity we are accustomed to seeing in causal talk about absences.) USE GOOD REPRESENTATIVES might explain these cases as well: for example, 'absence' expressions might represent merely possible causal paths.

we ascribe causal responsibility not to particular events but to causal paths, and a bit about the sense in which some but not all events on a causal path are "good representatives" of that path. First, in saying '*c* was the cause of *e*' or '*c* was causally responsible for *e*' we almost never mean that *c* alone was causally responsible for *e*. We know, after all, that the other events on the causal paths through *c* to *e* are partly responsible for bringing about *e*, in the sense that *e* would have been prevented, or would have happened in a different way, if any of many of the events on the causal paths through *c* to *e* had been prevented or had occurred differently than they did. But for many purposes little or nothing would be gained by mentioning multiple events on a causal path to *e*. Often, surprisingly few "landmarks" suffice to enable us to pick out the features of represented causal paths that are relevant to a particular conversation. So it is often most efficient to pick representatives of causal paths in conformity with the following constraint: they should provide conversational participants with enough information about the represented causal paths for the purposes at hand. And (as ever) speakers should mind the costs of adding misleading or unnecessary information to a conversation. Ceteris paribus, the better an event does at satisfying this constraint, the better a representative it is. Of course, when predictable inferences would lead the addressee to importantly false conclusions, or when important features of the causal paths are obscure or surprising, the speaker may use more representatives (or other devices) to ensure that her addressee draws out the relevant features of a path with appropriate, accurate detail. But very often a speaker can achieve her conversational goals by picking out a single apt representative for some causal paths—often, an event that has moral significance, or could have been easily prevented, or is such that a change in it, holding certain aspects of the situation fixed, would have made differences relevant to the conversational participants.

Even given all this, one might worry that inevitably too many or too few things will count as "good representatives." More concretely, the fact that (2) and (3) are felicitous in many contexts shows that both the leak and the puddle can be good representatives of the causal paths through them to the ice. Because USE GOOD REPRESENTATIVES does not demand that we pick exactly one event to represent a causal path or paths, it might look powerless to explain the infelicity of (7) and (8). But this is not so. As our judgments about (2) and (3) show, we start in a context in which both the leak and the puddle are good representatives of the causal paths through the leak to the ice. For precisely this reason, we do not generally need to use both the leak and the puddle to represent paths through the

leak to the ice well enough to satisfy USE GOOD REPRESENTATIVES. Once the leak is used to represent those paths, if it does so well enough to satisfy USE GOOD REPRESENTATIVES, the puddle becomes a poor representative of those same causal paths. To cite the puddle, too, as their representative would be to add information to the conversation that has been rendered unnecessary enough to be misleading. Using the leak to represent the relevant causal paths makes otiose—mid-utterance—the representative role that the puddle otherwise could have played.

USE GOOD REPRESENTATIVES also explains why it is easy to hear (7) and (8) as describing a situation in which neither the leak nor the puddle is causally relevant to the other. This is because when a speaker uses both the leak and the puddle as representatives of some causal paths, she conveys that both the leak and the puddle are needed to successfully represent some causal paths—that neither plays a representative role successfully discharged by the other alone. From this point it is natural for the addressee to hypothesize that the leak and the puddle play distinct representative roles by representing distinct causal paths—that because the puddle is needed to represent causal paths that are not well represented by the leak, the puddle represents paths that do not even include the leak.

Of course, the selection of a particular representative for a causal path or paths is not irrevocable. But it is difficult to switch from one representative to another. This is similar to the difficulty in switching between particular restrictions on domains of quantification. Speakers regularly use 'everyone' not to talk about absolutely everyone but rather about every employee, or every child, or every addressee, or what have you. But once a particular restriction on 'everyone' is in place, considerable conversational pressure holds it there. The reason for this is that it takes group effort for conversational participants to coordinate on domains of quantification, and one should not ask conversational participants to engage in such an effort without good reason. Similarly, it takes group effort to coordinate on a new representative for a given causal path. The bar to switching between representatives means that to pick one causal representative is in effect to screen off other potential representatives to some degree, making it important to pick representatives of causal paths that will well serve both present and (perhaps unconceived) future conversational purposes. For all a participant in a typical conversation knows, future conversational purposes could well demand especially good representatives. So there is pressure to select representatives for a path that are not significantly bettered by other potential representatives—a pressure counterbalanced, as always, by the demands of conversational and cognitive economy.

One prima facie problem for my use of USE GOOD REPRESENTATIVES is that it is easy to imagine sentences like (9)–(11) being felicitous descriptions of the situation we have been considering:

(9) The leak was a cause of the ice by being a cause of the puddle.
(10) The leak was a cause of the puddle, and because of that was a cause of the ice.
(11) The leak was a cause of the puddle, and the puddle was a cause of the ice, so the leak was a cause of the ice.

For example, we might use these sentences to explain why, given that we can fix neither the drain nor the heater, we ought to fix the roof. These sentences explicitly allot two representatives—the leak and the puddle—to the causal paths through the leak through the puddle to the ice. But this does not undermine my explanation of the infelicity of (7) and (8). Why might we allot two representatives to these paths? Consider some contexts in which we might actually use (9)–(11). Often, we use such sentences to explain why the leak was causally relevant to the ice, in speaking with someone who grants that the puddle was causally relevant but does not believe that the leak was, too. The causal paths that we are trying to inform our addressee about need (for these purposes) to be represented both by the leak and the puddle, for—as far as the addressee knows—the paths through the puddle to the ice do not include the leak. We inform addressees that some of those paths do include the leak by being explicit about the fact that we are using both the leak and the puddle to represent those paths.

Here is another situation that illustrates how USE GOOD REPRESENTATIVES can do helpful explanatory work. Suppose that the increasing heat caused both the expansion of the gas and the weakening of the container, and that the expansion and the weakening in turn jointly caused the rupture of the container. A fully informed speaker may appropriately describe this situation using any of (12)–(15).

(12) The expansion of the gas was a cause of the rupture of the container.
(13) The weakening of the container was a cause of the rupture.
(14) The increasing heat was a cause of the rupture of the container.
(15) The expansion of the gas was a cause of the rupture of the container. The weakening of the container was also a cause of the rupture.

But for many purposes (16) and (17) are sub-par (indeed, misleading) descriptions of the case.

(16) The expansion of the gas was a cause of the rupture of the container. The increasing heat was also a cause of the rupture.
(17) The weakening of the container was a cause of its rupture. The increasing heat was also a cause of the rupture.

For some purposes the increasing heat is a good representative of the causal paths through it to the rupture of the container. For other purposes, the expansion of the gas and the weakening of the container are good representatives of those paths. In this quite symmetric case, our purposes are often of this general type—they are purposes for which either the relatively 'distal' increasing heat or the relatively 'proximal' expansion of the gas and weakening of the container are crucial. But if the weakening was overdetermined—if it would have occurred whether or not the heat had increased—(17) is likelier to sound felicitous. This is because the weakening would not counterfactually depend on the increasing heat, making the heat a somewhat misleading (and thus somewhat flawed) potential representative of the causal paths through the weakening to the rupture.

As I mentioned earlier, my goal here is not to fully characterize the notion of goodness at play in USE GOOD REPRESENTATIVES. I want to use USE GOOD REPRESENTATIVES only to help structure our explanations of how conversational context influences the assertibility conditions of causal claims. Why does this structure matter? Consider two questions:

1. Which events that are causally relevant to *e* can felicitously count as a cause of *e*, in which contexts?
2. Given a particular causal path to *e*, which events can felicitously represent that path, in which contexts?

Both these questions are very difficult, and both have (to say the least) a significant empirical component. We are nowhere near to having systematic answers to either of them. But the second question is less difficult than the first. Ceteris paribus, it is much easier to compare how well two events that share a causal path to *e* do at representing that path, in a context, than it is to compare how well events on different causal paths do at counting as causes of *e*. As we will see, in certain important cases it is easy to see how various events fare in the competition to represent a given causal path.

II. TRANSITIVITY AND OVERGENERATION

A common move in the literature on causation, events, and causal transitivity is to argue that a theory overgenerates causes: it offends common sense by counting some event (or fact, or whatever) as a cause of another, when, intuitively, it is not. And so the theory is deemed false, or at least 'costly'.

In a sense we ought to worry about this kind of objection only to the extent that we aim to say what it is for one event to count as a cause of another in a particular conversational context. If our aim

is simply to say what it is for one event to count as causally relevant to another, then the fact that some events that are causally relevant to *e* rarely if ever count as a cause of *e* may be utterly unsurprising. Consider Lewis's suggestion that his analysis of causation is really an analysis of "causal histories," parts of which "will not be at all salient in any likely context...: the availability of petrol, the birth of the driver's paternal grandmother, the building of the fatal road, the position and velocity of the car a split second before the impact."[8] If, with Lewis, we hypothesize that causation is transitive, then we can expand this list indefinitely. One part of the causal history of the building of the fatal road, for example, is the crew's painting of the north curb line. So the crew's painting of the north curb line is part of the causal history of the accident. Granted, it rarely if ever counts as a cause of the accident. But (the arch proto-contextualist might say) so what?[9] We as metaphysicians are interested in the metaphysics of the causal relevance relation—a metaphysics that would yield sufficient conditions for the assertibility of sentences of the form '*c* was a cause of *e*' only if it were supplemented with a sufficiently comprehensive semantics and pragmatics.

Whether one thinks this gambit is principled will depend, I think, on what one thinks of notions like causal relevance and Lewis's "causal history." To the extent that one thinks the philosophically interesting questions about causation are about a "broad and nondiscriminatory" causal relation[10] one may welcome the move to '*c* was causally relevant to *e*'. Philosophers who suspect that we have little helpful pretheoretic grip on a broad, inclusive causal relation may be assuaged by seeing detailed explanations of why a particular event is a relatively bad representative of causal paths to *e*, even supposing that it is causally relevant to *e*.

Consider

SWITCH

> A train departs Mountain Station. It comes to a fork in the track, where an engineer flips a switch so that the train continues on the right-hand track. The right-hand and left-hand tracks rejoin before reaching Valley Station. The train arrives at Valley Station, and the switching made no difference to the time and manner of its arrival.

[8] Lewis, "Causal Explanation," pp. 215–16.

[9] Lewis makes similar maneuvers in his "Scorekeeping in a Language Game," in *Philosophical Papers*, vol. 1 (New York: Oxford, 1986), pp. 233–49 [Originally published in *Journal of Philosophical Logic*, VIII, 1 (January 1979): 339–59]; and his "Elusive Knowledge," in *Papers in Metaphysics and Epistemology* (New York: Cambridge, 1996), pp. 418–46 [Originally published in *Australasian Journal of Philosophy*, LXXIV, 4 (December 1996): 549–67].

[10] Lewis, "Causation," p. 559.

Ned Hall asks: "Is [the engineer's] flipping the switch a cause of the train's arrival? Yes, it is, though the opposing reaction surely tempts."[11] Hall gives a number of reasons to think the flip of the switch is a cause of the train's arrival—reasons that perhaps should be construed as showing only that the switch was causally relevant to the arrival. But we also want to know why the opposing reaction is so tempting.

USE GOOD REPRESENTATIVES helps us answer this question. By saying

(18) # The engineer's flipping the switch was a cause of the train's arrival.

a speaker would, by USE GOOD REPRESENTATIVES, convey that the flipping of the switch is a good representative of the causal paths running through it to the train's arrival. But using the switch to represent those paths gives a misleading picture of which events on the paths matter most for normal purposes. It is at least natural to infer, from the claim that *c* was a cause of *e*, that whether *c* occurred made some difference to the likelihood of *e* occurring, despite the fact that the counterfactual dependence of *e* on *c* is not necessary for *c*'s being causally relevant to *e*. It is plausible that this is because we are often more interested in counterfactual dependence than in causal dependence. We want to know what would have made a difference; we want to know what consequences changes upstream would have had. (To be sure, causal dependence is often more prominent when ethical issues are relevant.) Events on the causal paths to the train's arrival on which the arrival does counterfactually depend do better, in this quite important respect at least, at representing those causal paths. And the flipping of the switch does not better the train's departure (say) in the other respects that might be relevant to the goodness of a representative. Thus the departure, among other events, is for many purposes a significantly better representative of the relevant paths than the flip is. So is the train's motion after the tracks rejoin. And if the engineer's flipping the switch were used to represent those paths, it would fill the representative role that the train's departure would better fill, thereby linguistically screening off the train's departure and other events that are better representatives of the relevant paths.

The problem is not that there is something wrong with screening off per se—it happens all the time. The problem, rather, is that many of the screened off candidates do better at representing the causal paths to the arrival than the flip. And as I argued earlier, it is bad form to choose a representative for a causal path if that representative

[11] Hall, "Causation and the Price of Transitivity," this JOURNAL, XCVII, 4 (April 2000): 198–222, see p. 205.

would serve our evolving conversational purposes significantly worse than other candidates would. All this is consistent with the supposition that the flip is causally relevant to the arrival. It also does not mean that the flip is a bad representative of causal paths through it which end in events that occur while the train is on the right-hand track. Indeed, it is plausible that the flip will be a good representative of those causal paths, since those events do counterfactually depend on the flip.

This sort of explanation does not depend on there being a total preordering of the events that are causally relevant to *e*, in terms of their fitness for being counted as a cause of *e*. Indeed, I doubt that events can be ordered in this way. Is the spark a better or worse candidate to count as a cause of the fire than the presence of oxygen? Is my dropping the glass a better or worse candidate to count as a cause of its shattering than its fragility? Even relative to particular contexts such questions do not always have good answers. So there are many pairs of events that are both causally relevant to some event *e* but that are not comparable with respect to their fitness for being counted among the causes of *e*. What the explanation does rely on is the claim that given a particular causal path some events are better representatives of it than others. When we ask, of the causal paths that run through the flipping of the switch to the arrival of the train, whether some events better represent them than the flipping of the switch, the answer is 'yes': after all, there is the departure, the train's movement after the tracks rejoin, the entire course of the train before it arrives, and so on.

What goes into making an event a good representative of a causal path is an important and difficult question. Arguably, some philosophers of causation have already made significant progress on it, under a different mode of presentation. When they proffer analyses that do better than "broad and nondiscriminatory" analyses at matching our intuitions about what can count as a cause of what, they have taken themselves to be doing metaphysics. But a philosopher with a nondiscriminatory theory of causation also can construe such work as providing resources that help her say which events count as good representatives of a causal path. Because conversational context helps determine which events are good representatives of a causal path, every metaphysical theory of causation needs a 'good representative' theory to mediate between it and our ordinary language judgments. The question is just how much work that theory should do, and how much work the metaphysics should do. Even if a 'broad' metaphysics of causation needs a more ambitious theory of good representatives than a 'narrow' metaphysics, the need for this kind of supplementation does not obviously cut against the broad metaphysics. So the

advocate of a broad metaphysics can simply integrate techniques of a narrower putative 'metaphysics' of causation to work in her linguistic theory of good representatives.

I do not want to pursue this line further here, however, because I think it is illuminating to see how much work USE GOOD REPRESENTATIVES can do without being tailored to a particular metaphysics. Hartry Field's BOMB case, like SWITCH, nicely displays its potential.

BOMB

Billy plants a bomb in a room. Suzy comes into the room, notices the bomb, and flees. Suzy later has a checkup and is found to be in perfect health.

The presence of the bomb is a cause of Suzy's fleeing, and Suzy's fleeing is a cause of her perfect health the next day. But "the bomb is not a cause" of Suzy's health, making this a "counterexample to transitivity."[12] Or so the story goes.

It is very important to be clear about what needs to be explained. The presence of the bomb can in fact be cited as a cause of Suzy's health, as long as it is also counted as a cause of her fleeing the room, or of her not being present at the time of the explosion. For example: the presence of the bomb at time *t* caused Suzy to believe that an explosion was imminent. And given that an explosion occurred soon after *t*, Suzy should be glad that the bomb was present. Indeed, the presence of the bomb was a cause of her good health, because the presence of the bomb was a cause of her fleeing before the explosion. Of course, we are sketching causal paths to Suzy's health using multiple events, as is often necessary when the causal relevance of one such event is in question. We saw this earlier with (11):

(11) The leak was a cause of the puddle, and the puddle was a cause of the ice, so the leak was a cause of the ice.

So (19) is not particularly odd.

(19) The presence of the bomb was a cause of Suzy's fleeing, and (given that an explosion occurred) Suzy's fleeing was a cause of her good health, so the presence of the bomb was a cause of Suzy's good health.

Our real task, then, is to explain the oddness of

(20) # The presence of the bomb was a cause of Suzy's good health.

uttered with no further elaboration on the details of the case.

[12] Yablo, "Advertisement for a Sketch of an Outline of a Prototheory of Causation," in John Collins, Ned Hall, and L. A. Paul, eds., *Causation and Counterfactuals* (Cambridge: MIT, 2004), pp. 119–37, see p. 119.

In principle, the presence of the bomb might represent causal paths through the explosion, and it might represent paths through Suzy's fleeing. Paths of the first kind are irrelevant because Suzy's fleeing makes the explosion causally irrelevant to her health. (To be sure, the explosion would have been causally relevant to Suzy's injury, but for the fortunate fact that she fled the room and prevented herself from being injured.) And the presence of the bomb does not represent paths through Suzy's fleeing to her health very well compared to other events on those paths. For example, her fleeing the room does better at representing the causal paths through her fleeing the room to her subsequent health than the presence of the bomb would, because (again) counterfactual dependence matters so much to us. If Suzy had stayed in the room she would not have been healthy the next day; if the bomb had not been present there would have been no explosion, so whether or not the bomb had been present Suzy would have been healthy the next day. By uttering (20) on its own a speaker screens off potential representatives of some causal paths to Suzy's perfect health, among them her fleeing the room—a significantly better representative of the relevant causal paths. So in most contexts, her fleeing the room can count as a cause of her health, but the presence of the bomb cannot. Whether the presence of the bomb is causally relevant to Suzy's good health is of course not dependent on context, but whether the presence of the bomb counts as a cause of Suzy's good health does depend on context. Generally it does not, but when the speaker does not let the presence of the bomb screen off her fleeing the room (as in (19)) it may.

Bomb is an especially interesting case for present purposes because it elicits diverging intuitions about causal relevance. As a matter of fact, Hall thinks that the bomb is not causally relevant to Suzy's health (pers. comm.); Lewis thinks it is.[13] It is not surprising, I think, that there are many disagreements about what is causally relevant to what. The chief difficulty here is that natural language will not always help

[13] Cf. "Causation as Influence," p. 194. Lewis has at least two rationales here. First, the bomb's presence causes Suzy's health to be caused in one way rather than another, so the bomb's presence causally influences the causal history of Suzy's health. As a result it is part of the causal history of Suzy's health ("Causation as Influence," pp. 194–95). Second, if "whole" causal explanations are "the biggest chunk of explanatory information that is free from error" ("Causal Explanation," p. 218), then Suzy's fleeing is part of the causal explanation of her health. It would be obscure why her fleeing is explanatorily relevant if the bomb's presence was not also part of the causal explanation of her health. So the bomb's presence must be part of the explanation too. But if causal histories just are "whole" causal explanations (as Lewis sometimes suggests (as in "Causal Explanation," pp. 218–19)), then the causal history of Suzy's health must include the bomb's presence as well.

resolve such disagreements. Our judgments about sentences like (20) do not constrain the metaphysics of causation as directly as many have thought.

Lewis's line on BOMB is not vindicated, of course, by the fact that he can explain why we do not usually count the bomb as a cause of Suzy's health. But the existence of such an explanation supplies a principled way for him to agree with common sense that the bomb counts as a cause of Suzy's health in relatively few contexts, while still maintaining that the bomb is causally relevant to Suzy's health. So in response to a philosopher who claims that BOMB is a counterexample to causal transitivity simpliciter,[14] or that such cases are counterexamples to the conjunction of transitivity and the sufficiency of counterfactual dependence,[15] Lewis could observe that the planting of the bomb is not, in ordinary contexts, a good representative of the causal paths through it to Suzy's health. USE GOOD REPRESENTATIVES affords a metaphysically neutral way for any theorist of causation to explain away our linguistic judgments about cases like BOMB.

III. OVERDETERMINATION

Sometimes two otherwise innocuous looking causal claims seem odd when taken together. Earlier we considered circumstances and contexts in which (2) and (3) are both appropriate on their own, but (7) is odd.

(2) The leak was a cause of the ice. (…So if we fixed the leak…)
(3) The puddle was a cause of the ice. (…So if we fixed the drain…)
(7) # The leak was a cause of the ice, and the puddle was a cause of the ice.

I explained this oddness in terms of a context change in the midst of (7). The first conjunct of (7) changes the conversational context by making the puddle an otiose (and thus poor) representative of the paths through it to the ice. Evaluated relative to that context, the second conjunct of (7) is most naturally interpreted as representing causal paths to the ice that do not include the leak. (7) is inappropriate at least to the extent that it is inappropriate to suggest that the leak and puddle are both needed to represent different causal paths to the ice, for speakers make that suggestion when they assert (7).

Context shifts within conjunctions are not at all unusual. In particular, in many contexts sentences that could be felicitously asserted in

[14] Yablo, *op. cit.*, p. 119.
[15] Hall, "Two Concepts of Causation," in Collins, Hall, and Paul, eds., *Causation and Counterfactuals* (Cambridge: MIT, 2004), pp. 225–76, see pp. 246–48.

isolation would be infelicitous if conjoined or asserted together. Suppose for example that yesterday in the park I saw two dogs—one the largest I have seen in months, and the other the smallest I have seen in months. I tell you that I saw a dog in the park yesterday. I could then felicitously say either (21) or (22).

(21) The dog was the largest I've seen in months.
(22) The dog was the smallest I've seen in months.

But (unless it has been months since I have seen any other dogs) I could not felicitously say

(23) # The dog was the largest I've seen in months, and the dog was the smallest I've seen in months.

The first conjunct of (23) changes the context in a way that makes it inappropriate to say the second conjunct: it makes the larger dog significantly more salient than the smaller dog, thereby affecting the interpretation of the second occurrence of 'the dog'. This is obviously not in conflict with the fact that (22) can be used, on its own, to make the smaller dog more salient than the larger dog. So by appealing to context change we can hold onto our intuitive judgments about (21), (22), and (23) without questioning the soundness of (appropriately restricted) conjunction introduction.

Skeptical of the force of arguments from the threat of systematic overdetermination, Ted Sider writes:

> Should we say that *a baseball* caused a certain window to shatter? Or that *the parts of the ball* caused the window to shatter? Or that *the event of the ball's striking the window* caused the window to shatter? Or that *the fact that the ball struck the window* caused the window to shatter? Or something else? One wants to say *all* of these things! That is certainly the natural view.[16]

I agree with Sider that the natural view is that all of these things are fine to say on their own. Nevertheless their conjunction is undeniably odd:

(24) # The baseball caused the shattering, its parts caused the shattering, the event of its striking the window caused the shattering, and the fact that the ball struck the window caused the shattering.

(24) sounds odd for a by now familiar reason: it wrongly suggests that the baseball, its parts, and all the rest are on different causal paths to the shattering of the window. It wrongly suggests that, like soldiers on a firing squad, the baseball, its parts, and so on are causally separate

[16] Sider, "What's So Bad about Overdetermination?" *Philosophy and Phenomenological Research*, LXVII, 3 (November 2003): 719–26, see p. 720.

from each other in the sense that there are differences between what is causally relevant to them, between what they are causally relevant to, or both.

One might find overdetermination of the sort evoked by (24) to make for such "an ugly picture"[17] that we are obligated to reject at least some of its conjuncts. Merricks, for example, avoids systematic overdetermination in part by holding that "If the baseball exists, it does not cause the shattering of the window."[18] But because it is plausible to explain the oddness of (7) by appealing to a context change in its midst—rather than by rejecting (2) or (3)—we are amply justified in asking whether the oddness of (24) and sentences like it could be due to similar context changes. Unless he can show that there is no such context change in (24), Merricks simply is not licensed to conclude that we should reject any of its conjuncts.

On my own positive view about this case, there are numerically distinct sufficient causes of the shattering, in competition with each other only qua potential representatives of causal paths. Let a *causal cluster* be a set of causal relata E such that exactly the same causal relata are causally relevant to all the members of E, and all the members of E are causally relevant to exactly the same causal relata. We can define one causal cluster's being causally relevant to another in terms of causal relevance between their members (and we can rely on context to indicate which sense of 'causally relevant' is in play). One causal cluster is *causally relevant* to another just in case all the members of the first are causally relevant to all the members of the second. A (possibly infinite) sequence of causal clusters $\langle \ldots, E_{n-2}, E_{n-1}, E_n \rangle$ constitutes a *causal cluster path* to E_n just in case E_{n-1} is causally relevant to E_n, E_{n-2} is causally relevant to E_{n-1}, and so on. As before, we confine our attention to causal cluster paths that are *maximal* in the sense that they have no proper supersequences with the same terminal causal cluster. Accordingly, we have

STRENGTHENED USE GOOD REPRESENTATIVES

> When you ascribe some causal responsibility for E to a causal cluster path to E, use good representatives of that path for the purposes at hand.

We can account for the leak/puddle case with USE GOOD REPRESENTATIVES; STRENGTHENED USE GOOD REPRESENTATIVES is not necessary. But because the latter applies to singleton causal clusters, it can be used to explain any data explained by the former.

[17] Trenton Merricks, *Objects and Persons* (New York: Oxford, 2001), p. 67.
[18] *Ibid.*, p. 57.

If wholes, parts, events, facts, objects, and so on all exist, then even granting that they are distinct from each other it is plausible that many of them are members of the same causal clusters and so compete to represent the same causal cluster paths. It is plausible, for example, that exactly what is causally relevant to the baseball is causally relevant to the atoms arranged baseballwise, and that exactly what the baseball is causally relevant to is what the atoms arranged baseballwise are causally relevant to. This is so even granting that the baseball is distinct from the atoms arranged baseballwise, for it simply does not follow from the distinctness of *x* and *y* that *x* and *y* are causally relevant to different things, or that different things are causally relevant to *x* and *y*. So STRENGTHENED USE GOOD REPRESENTATIVES can also explain the oddness of (24). The form of the explanation is by now familiar: if we have already used the baseball's atoms to represent causal cluster paths to *E*, then we have made otiose the representative role that the baseball previously could have played.

Some philosophers think causal powers indicate what exists, in a way that might seem to make trouble for the explanation just offered. For example, Jaegwon Kim argues that "To be real…is to have causal powers; *to be real, new, and irreducible, therefore, must be to have new, irreducible causal powers.*"[19] This particular argument is invalid: to be uncommon, for example, is not necessarily to have uncommon causal powers.[20] Other methods might suffice to establish Kim's conclusion or something like it, however. Sydney Shoemaker also thinks that causal powers play an individuating role, contending that

> …[W]hat makes a property the property it is, what determines its identity, is its potential for contributing to the causal powers of the things that have it. This means, among other things, that if under all possible circumstances properties *X* and *Y* make the same contribution to the causal powers of the things that have them, *X* and *Y* are the same property.[21]

I doubt that views like these make any trouble for my explanation of (24)'s oddness. By distinguishing *x*'s causal susceptibilities from what is causally relevant to *x*, and distinguishing *x*'s causal powers from what *x* is causally relevant to, we can say that *x* and *y* may have distinct causal susceptibilities and powers even if they are members of the same causal cluster. One causal susceptibility of alcohol, for example,

[19] Kim, "The Nonreductivist's Troubles with Mental Causation," in John Heil and Alfred Mele, eds., *Mental Causation* (New York: Oxford, 1993), pp. 189–210, see p. 204.

[20] Moreover, one might think Kim should allow that some "real" things are distinct in virtue of differences in their causal susceptibilities but not their causal powers.

[21] Shoemaker, "Causality and Properties" in *Identity, Cause, and Mind* (New York: Cambridge, 1980), pp. 206–33, see p. 212.

is its tendency to burn when exposed to open flame in the presence of oxygen. One causal power of alcohol is its ability to dissolve shellac. A quantity of alcohol has this causal susceptibility and causal power whether or not it ever actually burns as a result of exposure to open flame or is ever actually causally relevant to any dissolvings of shellac. More generally, membership in the same causal cluster is entirely a matter of what is causally relevant to what in the actual world. Causal susceptibilities and causal powers are not.

The importance of these distinctions is especially clear when we consider some intuitive differences between mental events and correlated physical events. Suppose the pain and the C-fiber firing are part of the same causal cluster. There might nevertheless be a difference between what would have happened if the pain had occurred without the C-fiber firing and what would have happened if the C-fiber firing had occurred without the pain, and such a difference could well reflect differences between the causal powers of pain and the causal powers of C-fiber firing. There are many ways to flesh out the details here. Suppose for sake of argument that Lewis's original account of causation,[22] reconstrued as an account of causal relevance, is right. Since Stalnaker and Lewis,[23] the most plausible semantics for counterfactuals have not validated antecedent strengthening. That is, those semantics allow for cases in which

1. If there were to be a C-fiber firing, the man would wince, and if there were not to be a C-fiber firing, the man would not wince. (So C-fiber firing, if it occurs, counts as a cause of the man's wincing on Lewis's original account.[24])
2. If there were to be a pain, the man would wince, and if there were not to be a pain, the man would not wince. (So pain, if it occurs, counts as a cause of the man's wincing on Lewis's original account.[25])
3. If the neurological basis of pain had not been C-fiber firing and there were to be a C-fiber firing, the man would not wince, and if the neurological basis of pain had not been C-fiber firing and there were not to be a C-fiber firing, the man would not wince.
4. If the neurological basis of pain had not been C-fiber firing and there were to be a pain, the man would wince, and if the neurological basis of pain had not been C-fiber firing and there were not to be a pain, the man would not wince.

[22] Lewis, "Causation."

[23] Lewis, *Counterfactuals* (Malden, MA: Blackwell, 1973) and Stalnaker, "A Theory of Conditionals," in William L. Harper, Robert Stalnaker, and Glenn Pearce, eds., *Ifs: Conditionals, Belief, Decision, Chance, and Time* (Dordrecht: D. Reidel, 1968), pp. 41–55.

[24] Lewis, "Causation."

[25] *Ibid.*

In many such cases, if a C-fiber firing and pain actually occur they will be members of the same causal cluster. Nevertheless, they will not have the same causal powers: even if it were to have a very different neurological basis from what it actually has, a pain could still cause a wince. So there is ample room for a nonreductive physicalist to hold that a mental event is a member of some causal cluster that also contains a physical event, while holding that some mental events have causal powers distinct from the causal powers of any physical event.

IV. CONTEXT SENSITIVITY AND THE METHODOLOGY OF METAPHYSICS

We started by looking for an illuminating story about 'causation itself'—a metaphysical natural kind the character of which we took to be independent of us, independent of the contingencies of our causal talk, and independent of our concept or concepts of causation. I argued that our ordinary causal claims are influenced by conversational context in significant ways, and so we turned to a not quite ordinary term—'causal relevance'—in the hope that it denotes the fundamental causal relation in a context-free way. But then we found that in certain cases philosophers disagree about what even counts as causally relevant to what—let alone which is the true theory of causal relevance. And I argued, further, that in some cases judgments about ordinary causal talk do not indicate what counts as causally relevant to what, because we can explain those judgments in principled but metaphysically neutral ways. In effect, I showed that once we begin to flesh out the theory that interfaces between our metaphysics of causation and our causal talk, there is a strong possibility that our causal talk will grossly underdetermine the metaphysics of causal relevance.

We should not be surprised to arrive at this point. Given that we need complementary theories—one metaphysical, one linguistic—to get substantive predictions about our causal talk, it would be fortuitous if our causal talk came anywhere close to uniquely determining the metaphysics of causation. In some respects this underdetermination is liberating, because we no longer have to worry about pesky cases like BOMB. But it also raises troubling methodological questions: if linguistic judgments dramatically underdetermine our metaphysics of causation, exactly what will help determine it?

For the time being I think we are justified in putting such skeptical worries to the side. Having a better sense of what work can be done by our linguistic theory provides us a better sense of where to look for examples that are genuinely probative for metaphysics. That said, whether a linguistic judgment is probative for metaphysics depends on the specific ways in which we explain the context sensitivity of causal

talk. So we cannot make much real progress on the metaphysics of causal relevance without better understanding causal talk. In addition, some of the overlooked features of causal talk raise new questions for metaphysics. For example, I think we should investigate the possibility that causes are best thought of not as particular events (or facts, or whatever) but rather as causal paths themselves. (Note that taking causal paths to be the causal relata is not necessarily to hold a "process" view of causation: one need not be committed to "conserved quantities" in the sense of Dowe[26] and Salmon.[27]) These are roughly instrumental reasons to be interested in the context sensitivity of causal talk: understanding it may help us hone the knife with which we try to carve causal reality at its joints.

But I want to warn against the thought that the study of causal talk is just instrumentally important—that we would do well to ignore it if we could only find a way to theorize about causal relations without the intermediary of judgments about causal claims. For an analogy, consider your initial, unarticulated philosophical curiosity about the nature of friendship. The property of being a friendly acquaintance of *P* is broad and nondiscriminatory, instantiated at least by anyone who in some conversational context counts as a friend of *P*. We can know quite a lot about this relation without knowing anything about the ways in which 'is a friend of' is sensitive to conversational context. But saying what it is for two people to be friendly acquaintances obviously does little to address our curiosity about friendship. Ignoring context sensitivity makes our task easier—it is clearly easier to satisfy our philosophical curiosity about friendly-acquaintance-ship than it is to satisfy our philosophical curiosity about friendship—but ease of theorizing does not warrant such a change in subject. Any philosophically respectable course here will have to deal with or work around the context sensitivity of 'is a friend of'.

Similarly, we cannot ignore the context sensitivity of causal talk without neglecting much that is of philosophical interest. This is because our unarticulated curiosity about causation is in part a curiosity about causal thinking, which is crucial to folk psychology, moral judgment, scientific reasoning, and a host of other philosophically rich topics. It is impossible to cleanly excise ordinary causal talk from ordinary causal thinking, so to study one is to study the other. And we should welcome this connection: to take just one

[26] Phil Dowe, "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory," *Philosophy of Science*, LIX, 2 (June 1992): 195–216.

[27] Wesley C. Salmon, "Causality without Counterfactuals," *Philosophy of Science*, LXI, 2 (June 1994): 297–312.

example, progress on the theory of good representatives would likely help us better understand the connections between normative and causal judgments. It is rarely cheering to see that we cannot make progress on a family of philosophical problems without better understanding some related conversational context sensitivity. But we should not assume that we can satisfy our initial curiosity about causation without such an understanding.

ERIC SWANSON

University of Michigan

CLOSURE ON SKEPTICISM*

It is received wisdom that the skeptic has a devastating line of argument in the following. You probably think, he says, that you know that you have hands. But if you knew that you had hands, then you would also know that you were not a brain in a vat, a brain suspended in fluid with electrodes feeding you perfectly coordinated impressions that are generated by a supercomputer, of a world that looks and moves just like this one. You would know you were not in this state if you knew you had hands, since having hands implies you are no brain in a vat. You obviously do not know you are not a brain in a vat, though—you have no evidence that would distinguish that state from the normal one you think you are in. Therefore, by modus tollens, you do not know you have hands. At least, the skeptic has a devastating argument, it is thought, if we grant him closure of knowledge under known implication, which many of us are inclined to do: roughly, if you know p, and you know that p implies q, then you know q.[1]

To say that this is an intuitively compelling argument is an understatement; the project of finding a reply that is not table-thumping, or obfuscating, or special pleading has exercised philosophers for some time. The steps of the argument have been scoured in detail to find cracks that will yield under pressure. Some of these efforts have been intriguing, and illuminating, and some, I think, even provide dialectical victories that shift the burden of proof back to the skeptic. For all this, though, as I will argue, we have missed a very simple point: though the skeptical argument above is valid, it has a false premise, namely, the claim that the thing we seem obviously to know *implies* the thing we seem on inspection obviously not to know. I will argue that this part of the argument cannot be repaired in a way that preserves the skeptical threat. Thus, if the skeptic wants to convince us to worry about our ordinary knowledge, he will have to come up with a completely different argument.

Closure of knowledge under known implication (hereafter "closure"), is necessary for the skeptical argument presented above but obviously not sufficient. For the closure principle to apply to our case, we would have to know that having hands implies that one is not a

* Thanks to John MacFarlane and Paolo Mancosu for helpful discussion.

[1] For an up-to-date discussion of this argument, see John Greco, "External World Skepticism," *Philosophy Compass*, II, 4 (July 2007): 625–49.

0022-362X/10/0705/243–256

brain in a vat. We cannot know that, as epistemologists are already aware, because the implication does not hold and false claims cannot be known. The implication does not hold because one could be a brain in a vat, so far as that is described above, with hands. The hands would be attached seamlessly to the brain, hence yours in an undeniable sense. These stipulations describe a scenario no less plausible than the original one of a brain in a vat. The scenario ruins the implication the skeptic needs because a handed brain in a vat is a counterexample to the claim that having a hand implies you are not a brain in a vat.

The Addams Family–Season 1 Courtesy of MGM CLIP+STILL

Epistemologists are aware that the implication claim first stated by the skeptic does not hold, due to the possibility just described, so the implication claim typically gets propped up in the obvious way, by saying that having hands implies one is not a *handless* brain in a vat. Sometimes one puts a tone on the emphasized word to convey the judgment that this detail is tiresome. One then moves along in development of the skeptical line to get to the more interesting issues, confident that the patch has done no harm to the argument because implication has been achieved. However, it is not enough that there be an implication. It must be an implication from something we think we do know to something we pretty clearly do not, in order to set us up for a modus tollens. What is wrong with this particular patch is

that weakening the conclusion to "I am not a handless brain in a vat" trivializes it for this purpose. If we assume I know that I have a hand, then we should not have the slightest hesitation to credit me with knowledge that I am not a handless brain in a vat.

No appeal to the closure principle is needed to support this conclusion. The claim is independently obvious because that you are not a handless brain in a vat is just not much to know. If we know that someone has hands then it follows that she is not a handless person with high blood pressure, or a handless victim of child abuse, but this would not give us any assurance that she need not go to a doctor for these conditions. To a person who already knows she has hands these claims say nothing at all about how far she might or might not be susceptible to heart disease or suicide. For this reason they are statements that it is trivially easy to know if you know that you have hands. If I know that I have hands, then in virtue of that I know I am not a handless anything. The implication is achieved in the skeptical argument, but only by letting the issue of brains in vats swing free of it.

The problem with my claim, one might think, is that it assumes that whether or not one has a hand is independent of whether or not one is a brain in a vat. The blood pressure example would look very different if not having a hand was correlated with having high blood pressure and you knew it. Then, indeed, finding you have a hand would give you a reason not to worry about your blood pressure. In our case, one might say, not having a hand is part of what we *meant* by being a brain in a vat. It is not an extra piece of news. The word "handless" gets added to the conclusion of the skeptic's argument only in order to make this explicit, so that one can see how clear the implication is. This idea is also a good explanation of our tone of tiresomeness—it should be obvious that a brain in a vat, in the sense we had in mind, has no hands. The implication holds, and the conclusion is not trivial.

If this is what we meant, then, I submit, it is not what we wanted to mean, or should have meant, given our collective state of puzzlement and distress over this skeptical argument, for the conclusion imagined is still trivial. Having a hand does make you distinct from the brain in a vat of imagination that has no limbs, but it does so in only one respect. It tells us nothing about whether you resemble it or not in any other respect. Let a brain in a vat be a thing that by definition has no hands. Having a hand still allows you to be a thing that is like a brain in a vat in every respect except that it has a hand seamlessly attached to it. The question now is how significant it is to find out that you are not a brain in a vat, when you still could be the same thing but for a hand attached; you still could be systematically deceived about just

about everything. The possession of these hands does not imply the thing we seem obviously not to know, which is that we are not subject to systematic deception. It is the latter concept that insures the intuition that we do not know the conclusion of the skeptical argument. Either the conclusion of the skeptic's argument is weak enough to be implied by the premise that I have a hand, but not strong enough to seem hard to know, or the conclusion of his argument is strong enough to appear obviously unknown to us, but not weak enough to be implied by my having a hand.

If I am right, then why have we been under the impression all this time that the adjusted conclusion "I am not a handless brain in a vat" is nontrivial? One reason is that philosophers are like all human beings in being susceptible to associational "thinking," that is, in drawing conclusions that have not been stated, purely on the basis of the proximity of words to one another. All people are sometimes victims, for example, of the devices of highly trained advertising agencies that do psychological research on how we are moved by associations. There was an ad recently that said, above a vivid picture of a train, "Legally, we can't say you can throw it under a train," of the TOUGHBOOK laptop computer. The ad did not assert that you can throw it under a train (and have it survive), but because precisely that clause was inscribed—see the original sentence—an exaggerated impression was created, in just about everyone I would venture, of just how tough the TOUGHBOOK is. Similarly, the words of our adjusted conclusion are "I am not a...brain in a vat," and this created a strong impression that this sentence without the ellipses had been asserted, or at least that some information was conveyed about this matter. Philosophers are not immune to such unconscious mistakes; we are all apt to make them when our conscious attention is directed elsewhere.

A second reason that the sentence "I am not a handless brain in a vat" seemed to carry the content that I am not a brain in a vat is conversational implicature. Suppose a man says that he enjoys talking to me. I ask him whether he has a wife and he replies "I don't have a wife I can *talk* to," where the word "talk" is not only emphasized but raised in pitch. The content of his reply contains no information about whether he has a wife. However, the emphasis conveys very clearly that he does. What is relevant about this case is that the content of the sentence is perfectly consistent with the message that he does have a wife, despite the fact that the sentence contains the phrase "I don't have a wife." Similarly, the content of the sentence "I am not a handless brain in a vat" is perfectly consistent with my being a brain in a vat. This is why it is even possible to make a strong suggestion that I *am* a brain in a vat, by saying "I'm not a *handless* brain in a vat," if

the word "handless" is emphasized and higher in pitch. To say the sentence "I'm not a handless brain in a vat" with a high-pitched emphasis on "handless" would reveal the triviality of the claim with respect to the matter of whether one is a brain in a vat (on the assumption one does know one has a hand), but I never hear epistemologists say the sentence that way.

The word "handless" is sometimes introduced with an emphasis that lowers the pitch on this word (to convey that tone of tiresomeness), but this hides the fact that no information has been conveyed that I am not a brain in a vat, just as "I don't have a wife I can talk to," may well fail to set off the wife alarm if the word "talk" is not raised in pitch. Admittedly, epistemologists also sometimes say the conclusion of our argument straight. In that case one is likely presuming that the word "handless" merely brings out an assumption already in what we meant by the phrase "brain in a vat," and politely leaving out the tone of tiresomeness. As I argued above, tone or no tone, the conclusion that follows is thereby trivialized, and easy to know if you know you have a hand.

Another plausible reason for the mistake is an equivocation on the term "brain in a vat." One could mean by this phrase a literal, specific image of a brain with no limbs or funny stuff, or one could mean this image as a kind of stand-in for any of a host of scenarios in which one is systematically deceived. Knowing you have a hand is plenty good enough to rule out the first, and miserably inadequate for ruling out the second, even if the host of scenarios is a set of small variations on a single theme. Our confidence in the implication has come from the first reading of "brain in a vat," and our confidence that the conclusion is something we do not know comes from the second. By equivocation we conclude that something we obviously do not know is implied by something we obviously do.

The initial patch I have described is of course not the only recourse the skeptic has. He could find a different way to weaken the conclusion, in which case the task is still to avoid making it trivially knowable. I will canvas another way of using this conclusion-weakening strategy below. The other obvious approach is to strengthen the premises. In this strategy we would keep the conclusion the same—I am not a brain in a vat in the originally intended sense—and add premises to make sure that what we think we obviously know does imply this conclusion it seems we clearly do not know. This turns out to be harder than it may seem, for even if we added claims that we have feet, and likewise for other body parts, things we know just as obviously as we know about our hands, the possibility of systematic deception does not go away. We could imagine an entity like a brain in a vat in every respect except that it had hands, feet, and so on, attached.

The number of attachments is not the issue in how much it takes to rule out systematic deception. The poor captured people who are used as batteries by the Matrix of movie fame have kept their entire bodies, but their brains are being fed impressions of a colorful world nothing like the dank storage facility in which their pods are suspended. This scenario would be as disturbing as the image of ourselves as "mere" brains in vats, and as obviously difficult to rule out. What makes something a brain in a vat in the relevant sense is that you are not related to the real world in the way you appear to yourself to be, and you have no indication of that; thus the world your hands and feet exist in is nothing like the world of your impressions. I will call this scenario in which you are systematically deceived one where you are a *brain in a vat* to indicate that this feature is essential to the scenario, while failing to have limbs, for example, is not. The denial of this envattedness, which it seems independently obvious we do not know, needs to follow from things we think we clearly do know; knowing that one has ever so many hands and feet does not rule out the disturbing and indiscriminable, and hence essential, feature of the *brain-in-a-vat* hypothesis.

What would rule out the skeptical hypothesis? The *brain-in-a-vat* hypothesis implies something about the vat-brain person's relation to the world: her lack of discriminating evidence about not only the world but her situation in it. The skeptic's premise must rule out all logically possible ways of realizing this. Once we understand this another repair strategy suggests itself. Merely that I possess hands is not enough, but perhaps this is because that claim does not say that the hand is connected up to my impressions, and intentions to move, in the normal way that it is when I have evidence and a nondeceptive set of impressions of the world. It seems that the claim of a hand that is normal in the relevant way—which we can as innocently agree we have knowledge of, when the skeptic asks, as we can agree about the previous claim—will do the trick of implying the claim that I am not a *brain in a vat*, since the *brain-in-a-vat* hypothesis says I am systematically deceived about everything, and this says there is something about which I am not deceived.

Call the first type of hand that is unconnected to my impressions a "floppy" hand, and the second a "hooked-up" hand. The problem is that a hooked-up hand is also far too little to rule out the *brain-in-a-vat* hypothesis. Though the hands of the poor people in the actual Matrix are floppy in the sense just introduced, we can easily imagine them having hooked-up hands, as long as we also enlarge the pod to allow their free movement. Their movement would require some movement of the arms, but they have those too and we can imagine them

hooked up. The impressions they have of their hands and arms, both sensory and motor, would come from the hands and arms, whereas their impressions of everything else would come from the supercomputer stimulations. The real and the fake would have to be coordinated with each other, the fake impressions of objects responding just as real objects would, to the interventions of the real hands. But there is nothing impossible about this.

An instance of the idea would be a video game: your control of the joystick is real, but what it is controlling is representations of things that are not real, and what it is controlling is a world that the player can increasingly come to inhabit as if it is real. Suppose such a player becomes fully entranced, without any longer having a sense of the set-up or movement of the rest of his body. Then he is systematically deceived. He will not come out of that world by any prompt within the game-world, but only by a screen that pops up saying he has run out of money, or by a bout of thirst, or intervention from a parent. We can imagine a case in which none of those external cues are available. It is clear that having hooked-up hands does not imply that one is not a *brain in a vat* any more than having a collection of floppy hands and feet did. When we want to know that we are not systematically deceived we expect more than that there is one little thing from which we are not hopelessly unconnected. Thus the "not" in the phrase "I am not a *brain in a vat*" does not function intuitively as it seems it should logically, issuing in a weak claim because it is denying a strong claim. Intuitively, "I am not a *brain in a vat*" means that most ordinary things are pretty much as they seem. If it does not mean this, then it is a claim that is too easy to know for the skeptic's purposes—just wave your hand. The reason the denial of the real *brain-in-a-vat* hypothesis is so strong is that "I am a *brain in a vat*" is a disjunction of lots and lots of quite similar ways one could be uncorrectably deceived about just about everything: a brain disconnected from everything except your little toe, deceived about everything except the existence of the floor, and so on. Each of those is easy to know in virtue of its being easy to know you have a little toe, and that there is a floor, but no one or two of them denies the sort of systematic deception the *brain-in-a-vat* hypothesis is about.

One might wonder if the problem is that we have not taken into account enough body parts. Having a greater number of floppy body parts did not help, but maybe it will if the numerous parts are hooked up. However, we can draw out the scenario just discussed with any number of body parts we like by imagining the interface between the joystick and the hand growing into an interface between the entire body and a control surface. Now my whole body is doing every

motion I think it is doing, and I am feeling whatever is impinging on the surface of my body. However, none of my impressions bears any indicative relation to the way the world is. The body is pushing and pulling around a real interface, but the interface is pushing around false representations (from my point of view), or objects that do not match my impressions (from an objective point of view), or nothing at all. I could be a whole-body-hooked-up *brain in a vat*.

Perhaps, then, it is not about me and my body, and ruling out the *brain-in-a-vat* scenario requires adding to the premises some things that I apparently obviously know about the *world*. Take the table of skeptical lore. This will not do either, since that premise typically states only that a table exists, and we already know that mere existence allows the possibility of floppiness—where I have no appropriate connection to the table. What if we suppose that I am hooked up to the worldly object, the table, in some appropriate way, say causally. Suppose also that my visual impressions of the table are perfectly coordinated with my other impressions of the world, whether those are fraudulent or true. But this does no good. I could be resting my arms on a table while the rest of my being is perfectly engrossed in a video game on the screen in front of me. The table impressions are properly produced: it is not just that when I have the leaning feeling in my arms it is because my arms are leaning, as we had already with the hooked-up arms, but also that when the arms are truly leaning, and I am having the impression of their leaning on a table, they are leaning on a real table. Apart from the hooked-up table, though, the entire world of my impressions is a fraud. Make the screen bigger and bigger until it surrounds me; hook me up to a feeding tube; make sure the game world never ceases to be interesting and has a backup generator; imprison anyone who might care to save me; and I am a *brain in a vat*.

One might think that the problem is that we are only considering hooking me up in the normal way to *one* object. The world has many objects, and if we suppose me hooked up to many, many of them, then we are imagining a scenario in which I surely cannot be deceived about very much in my physical surroundings. Is that not enough? Given that we are assuming from the previous steps that I know I am hooked up to my whole body, too, does this scenario not, for all intents and purposes, rule out the possibility that I am a *brain in a vat*? We can look at this approach in two different ways, as either a strengthening of the premises or a weakening of the conclusion. We will see that neither strategy helps the skeptic.

Our strategy now will be to put into the premises enough claims about body parts and objects that a robust denial of the *brain-in-a-vat*

thesis will be implied. In strengthening the premises this way we want to include enough hooked-up objects to insure that I am pretty much in touch with the world around me. Throw in the table, the chairs, the kitchen sink, the lamps and couches, the truck I see outside the window, the sunshine, the floor and ceiling, the walls. Are we there yet? Does all of this imply we are not systematically deceived in the appropriate sense? One problem is that there are a whole lot of things left off of this list. Does the friend you think you just talked to on the phone exist? Is there really a building supporting the room you are sitting in when you are not looking at the building? Why think that closed closet door does not open into outer space? Assume that you do know all of those things you list. They do not imply what the skeptic needs because the list you make, however long, will always leave out an infinite number of important aspects of the world. On the other hand, the things you will manage to list do not appear to imply anything that it would be surprising to think you know on the assumption that you know *them*. For example, it will not be doubtful that you know your hand is not a fake hand. To assist the skeptic we constructed strengthened premises with the object of making them imply a denial of systematic deception, so we had to assume that the hand you know about is a hooked-up hand, a really hooked-up hand.

The list of things that would need to be claimed in a set of premises implying the denial that one is a *brain in a vat* is of course infinite. However, given infinite time one could verify each claim on the list, the way one does with the claim that one has a hand, by directly inspecting them seriatim. The problem is that the knowledge so produced that the closet door does not open to outer space expires when I move away to inspect the lamps in the living room. Can we not have that knowledge in a different way? Not if we are trying to help the skeptic, whose target is those of our beliefs that we think we most obviously do know. We need to make the claims that go into the skeptical argument's premises very, very hard to believe I do not know, the way that it is hard to believe that I do not know I have a hand, since I can feel it and wave it in front of myself. Much if not all of our confidence that we know we have hands is this direct verification. This cannot be done with all of the claims we need in the premises, even if we cut off the list to a large finite set, because we cannot sufficiently directly verify them all at the same time, even roughly, which is what we need to do in order to assert our knowledge of them as premises of a single argument.

The things that we can claim simultaneously obviously to know do not appear to be strong enough to imply anything that we obviously do not know, and so surely not that we are not systematically

deceived. But one might think there is an obvious solution to all of this. You can express all of that information, that there is a table, chairs, sunshine, a building supporting me, whatever you see, simply by making a generalization that includes all of those examples without listing them individually. The generalization captures everything we need in one expression, perhaps making it possible to verify it all at the same time. What would the generalization look like? In order to capture all the things that I should be properly connected to if I am going to rule out being systematically deceived I must say that, modulo correctible errors—false beliefs which observations potentially could correct—things are pretty much as they appear to me to be not just at this moment, but also according to the general assumptions that the perceptual process typically has me making, such as that objects do not disappear in virtue of my turning away, and so on. Thus, that there is a building holding up my office counts as part of how things appear to me to be in this sense of "appear."

But now we have come full circle. In order to get premises strong enough to imply the conclusion that I am not a *brain in a vat*, we have had to add so much information to the premises, and in such a generalized form, that if we know those premises, then there can be no surprise that we also know we are not *brains in vats*, for what is left for us to be systematically deceived about? We may be wrong about many things, even ordinary things, but only in the normal way of being wrong, not uncorrectably so. We have closed the implicational gap, but only by inflating the premises to the point of recognition. Alternatively, we might think that the strengthened premises of the argument are far too much for us to know in any obvious way. In that case they also give us no reason to think we know we are not *brains in vats*, but that does not give us a modus tollens since there is no assumed obvious knowledge for it to undermine.

It still may seem that we have something to worry about, in that we have exposed that we may not know that we are not *brains in vats*. Sure, we know we have hands, but what we see now is that even if we assume closure this does not mean we know we are not *brains in vats*, because that does not follow. We do, surely, go around implicitly believing we are not so thoroughly deceived, though, so if we cannot defend that claim there still seems to be a problem. Part of the reason for this worry is not yet having fully taken on board the claim of this paper. Lack of knowledge that you are not a *brain in a vat* undermines your claim to knowledge only of those things inconsistent with your being a *brain in a vat*. A given list of beliefs about things around us being thus and so, and even our being rightly hooked up for knowing

that they are thus and so, is obviously not inconsistent with being a *brain in a vat*. This may seem like a bad thing—all of the things we are most confident we know will never get us to the reassuring knowledge we are not otherwise systematically deceived about many, many things. But it is just as much a good thing: we do not need to know we are not *brains in vats* in order to know however long a list of the familiar things we think we know. For all the skeptic has done, we can take the skeptic's first premise—you know that you have hands—and go home with it. We can take our feet home too, and keep assuming we know the closet door does not open into outer space. Nothing in his subsequent argument touches what we are permitted to think we know of such things.

The kicker, one might think, is in those assumptions that perception has us automatically making, such as that objects remain when I am not looking at them. Such claims are generalizations and so not claims I can directly verify in the way discussed above, yet we believe them and think we know them. However, granting that we think we know such generalizations, and granting that we cannot verify them directly, this still does not pose a problem. The skeptic has not shown that direct verification is necessary for knowledge. We think of this standard because the skeptic focused on an example, the claim that we have hands, where we fulfill it, and he focused on this because direct verification seems of all things overwhelmingly sufficient for knowledge, and he needed a premise we seem very obviously to know. We did not need to assume that direct verification is necessary for knowledge in order to take the skeptic seriously in the first place, and his argument leaves the question whether we know these generalizations just as it was found.

The effect of the argument of this paper somewhat resembles the outcome of views of knowledge that deny closure. In both you have a split decision where it is possible for you to know you have hands without knowing you are not a *brain in a vat*. But here the reason for the split is that it is possible to be a *brain in a vat* even if you have hands. The difference is in whether we deny that knowing you have a hand, and knowing that your having a hand implies that you are not a *brain in a vat*, implies that you know you are not a brain in a vat (closure); or deny that "I have a hand," and claims relevantly like it, imply "I am not a *brain in a vat*." There is no need to deny closure in order to defeat the skeptic in the way advocated here. There is no need to deny any general principle about knowledge, as far as I can see. Here, we got generality over the moves the skeptic might make to repair his situation by explaining the trade-off he will always face in trying to identify both a logical implication and

a huge intuitive knowledge gap. The skeptic needs a conclusion strong enough to be obviously unknown by us, and weak enough to follow from something we obviously know. His problem is that the closer we get to an implication, the farther we get from this intuitive combination.

The argument of this paper clearly does not appeal to a denial of closure, but one might think it tends to suggest the opposite, closure, and even, perhaps, to depend on it. This is because it is sufficient for a counterexample to closure if we find a case where we obviously know something, obviously know that it implies something else, and obviously do not know the something else. If I am right that the skeptic cannot find the kind of example he needs then it looks like a counterexample to closure cannot be found either. This is not quite right. There may be counterexamples to closure that do not aid the *brain-in-a-vat* skeptic. For example, one might think that knowing that one has a hand does not give one knowledge that it is not a fake hand although the first implies the second. However, this will not help a *brain-in-a-vat* skeptic for if this is a failure of closure then one does, or can, know one has a hand despite not knowing that it is not a fake hand, and thus that one is not a *brain in a vat*. Thus, one does not get to do a modus tollens to undermine the claim to knowledge of a hand.[2] *Actual* counterexamples to closure do not undermine my argument. However, all of this is very confusing. The approach to defeating skepticism that denies closure[3] assumes the skeptic needs closure in order to make his argument go. The argument of this paper suggests that what the skeptic needs is a counterexample to closure. How could both of these be true of his one argument?

[2] The same goes for the potential counterexample from "That is a zebra" to "That is not a cleverly disguised mule." If knowledge is not closed then not knowing whether there is a subterfuge does not undermine your knowledge that it is a zebra. Of course, one might think of these hand and zebra examples as helpful to the skeptic because if one maintains closure then the apparent fact that we do not know the conclusions comes back to undermine our knowledge of the premises. My response to this is to ask exactly what kind of hand it was you thought you knew about when you claimed that you knew you had a hand. It seems that would have been a real hand, in which case why exactly do you not know it is not fake? The other possibility is that you were claiming you knew you had an at least floppy hand, in the sense above, in which case there is no reason to expect you would know it was not fake, and so no modus tollens. A similar point can be made about zebras and mules.

[3] See Fred Dretske's "Epistemic Operators," this JOURNAL, LXVII, 24 (December 1970): 1007–23, "Conclusive Reasons," *Australasian Journal of Philosophy*, XLIX, 1 (May 1971): 1–22, and *Knowledge and the Flow of Information* (Cambridge: MIT, 1981), and Robert Nozick's *Philosophical Explanations* (Cambridge: Harvard, 1981).

Closure and closure failure are both at work because the skeptic's argument is a reductio ad absurdum, and the way down is different from the way up. The way down appeals to an implication claim, and two intuitions:

(1) I know "I have hands."	Intuition
(2) "I have hands" implies "I am not a *brain in a vat*."	"Logic"
(3) A normal person and a *brain in a vat* have the same evidence.	definition, stipulation, or...?[4]
(4) I (obviously) do not know "I am not a *brain in a vat*."	(3) plus (independent) intuition

If all of these statements are true, then this is a counterexample to closure, in which case the skeptic fails to undermine our knowledge of our hands. This is not what the skeptic wants, but he does need to make these four statements all *look* true. He needs to produce an apparent counterexample to closure on the way down, but one that does not actually disturb your conviction that knowledge is closed, so that your only option is to do a modus tollens and lose confidence in your ordinary beliefs. Another way out would be Moore's dogmatic one, of course, insisting that because one knows one has a hand, and because knowledge is closed, therefore one does know that one is not a *brain in a vat*, but the skeptic hopes you find that laughable, and if my argument is right then Moore's argument also has a false implication claim.

The option of denying closure admits that (1)–(4) are all quite convincing and denies the skeptic the move from "I do not know 'I am not a *brain in a vat*'" to "I do not know 'I have hands'." In other words, it denies the way up. What is distinctive about the argument here is that I am denying the skeptic the way *down*, via an argument that his apparent counterexamples to closure are illusory, and intuitions to the contrary are due to sloppiness about implication. Once we see what we need for his implication claims we see by inspection, independently, that the premises we clearly know or do not know line up only with conclusions we clearly know or do not know, respectively; the skeptic has not created a problem or a reason to deny closure, because he has a problem defending (1), (2), and (4) simultaneously.

[4] In *Knowledge and Its Limits* (New York: Oxford, 2000), Timothy Williamson points out that the skeptic's argument is not serious if he plans simply to stipulate that the brain in a vat and I could have the same evidence. "Same evidence" must be defined, and the claim that it is possible defended, which Williamson argues cannot be done. Whether or not this challenge can be met, I grant the skeptic's assumption here for the sake of argument.

The argument here proceeds without Moorean-style dogmatism, for I do not claim that we do know we have hands or that we do know that we are not *brains in vats*, only that the skeptic has given us no reason to think we do not. I also differ from Moore in refusing the claim that knowledge of something momentous—that I am in large part undeceived about my body and the world—follows from knowledge of something skimpy, such as that I have one or two real hands. The knowledge does not follow because the thing itself does not follow. Too little attention had been paid to how much is necessary for an implication claim and what the contents of the imagined premises and conclusions were or must be if the goals of the argument were to be achieved.

We can defeat the skeptic without denying closure because in his initial foray he needs to convince us of an apparent violation of it but only comes up with a case where the principle does not apply or else is not intuitively violated. Anything you know as well as that you have a hand will carry so little information that it will not imply you are systematically undeceived about much of the world. Anything that is so informative as to imply this is either something we do not plausibly—and certainly do not obviously—know, or else something the knowing of which would also make us obviously know we are largely systematically undeceived. The kind of example the skeptic needs is a will-o'-the-wisp.

None of this means that we need to worry that we are knowledge-poor. It means that the skeptic's argument has not shown anything about our knowledge. Rather, he has engaged, with our assistance, in an iterated shell game. (Ten dollars if you can tell me where the knowledge went!) Most people think, *contra* G. E. Moore, that you cannot get out of radical skepticism by waving your hands. What we have seen here is that you cannot get *to* a radically skeptical challenge by hand-waving either. These are both true for the same reason: you should not expect knowledge that you have a hand to give you knowledge of a world, not because of closure failure but because there being a hand does not imply there is a world, much less one that is like we think it is. Even the Romantics, who told us that we can see the world in a grain of sand, or the universe in a drop of water, did not think we could expect to do so by logical implication.

SHERRILYN ROUSH

University of California, Berkeley

COMMENTS AND CRITICISM

WHY RESPONSIBLE BELIEF IS BLAMELESS BELIEF*

Let *Doxastic Deontologism* (DD) be the following thesis:

(DD) *S* is justified in believing that *p* iff *S* believes that *p* responsibly.

The idea behind DD is that we are under certain doxastic obligations and that we should understand the justification of belief in terms of our compliance with them.

But what is it to believe responsibly? That is a difficult question. Let us consider a couple of conflicting proposals that have been put forward recently, in order to (at least partially) elucidate the concept of responsible belief. Let us formulate the thesis that responsible belief should be cashed out in terms of praiseworthiness, the thesis of *Doxastic Deontologism as Praiseworthiness* (DDP), as follows:

(DDP) *S* responsibly believes that *p* iff *S* is praiseworthy for believing that *p*.

And let us distinguish DDP from the more standard deontological view of doxastic responsibility, which is cashed out in terms of blamelessness, a view we call *Doxastic Deontologism as Blamelessness* (DDB):

(DDB) *S* responsibly believes that *p* iff *S* is blameless for believing that *p*.

We argue that if DD is true, then DDB rather than DDP is true. We also attempt to refute some arguments in favor of DDP, mainly those recently proposed (or at least inspired) by Brian Weatherson.[1]

We will assume with Weatherson that DDP and DDB are mutually exclusive. It seems that if one is praiseworthy for believing that *p*, then one is also blameless for believing that *p*, but that if one is blameless for believing that *p*, one is not thereby praiseworthy for believing that

*We would like to thank Jeroen de Ridder, Herman Philipse, Brian Weatherson, and René van Woudenberg for their helpful comments on earlier versions of this article.

[1] Brian Weatherson, "Deontology and Descartes' Demon," this JOURNAL, CV, 9 (September 2008): 540–69. Page references will be to this article, unless otherwise indicated. Weatherson never gives a precise formulation of the view he advocates, but DDP seems to be a correct representation of his main thesis: "justification is a kind of praiseworthiness, and ... praise is more relevant to epistemic concepts than blame" (p. 540); "justification is a kind of praiseworthiness, not a kind of blamelessness" (p. 543); and "a belief's being justified is not a matter of it being blameless, but a matter of it being in a certain way praiseworthy" (p. 569).

0022-362X/10/0705/257–265

p. That is, in certain circumstances one is responsible[2] for holding some belief *B*, while being neither blameworthy nor praiseworthy for holding *B*. On DDB, we believe responsibly in those circumstances, whereas on DDP we do not. The views do not only exclude each other, they also seem to be the only games in town. Any other option, such as the view that responsible belief is blameworthy belief, seems absurd.

This paper is organized as follows. In sections I–III, we discuss three arguments putatively in favor of DDP. In section I we respond to the idea that DDP enables us to deal with the familiar problem of doxastic involuntarism. In section II we discuss the claim that DDP is necessary to account for all our intuitions in the New Evil Demon case. And in section III we consider an argument to the effect that DDB, in conjunction with three principles a doxastic deontologist is bound to accept, leads to a contradiction. We show that none of these arguments stand up to scrutiny.

I. PRAISEWORTHINESS AND CONTROL

Let us first deal with the idea that DDP can handle the voluntarism problem for deontologism. The problem is that, since we do not have voluntary control over our beliefs, and 'ought' implies 'can', we cannot be said to bear doxastic responsibility. However, it may be that praiseworthiness "outruns" voluntary control, so that DDP solves the problem. Consider Weatherson's cricket captain, who comes up with a particularly imaginative field placement during a match. While we may want to praise the captain, we do so despite the fact that he had no control over that. After all, claims Weatherson, coming up with the particular field placement is hardly something one can set out to do. And we deem this captain more praiseworthy than his colleague, who works equally hard, but does not come up with such an imaginative field placement. So if we can justly praise *S* for φ-ing, regardless of whether φ-ing is under *S*'s voluntary control, then DDP is immune from the voluntarism problem. However, for at least three reasons we do not think appealing to praiseworthiness helps deontologism here.

First, in order for the reply to succeed, being imaginative needs to be analogous to being in belief states, such that we have no *direct* voluntary control over either. However, we do seem to have *indirect* voluntary control over and indirect voluntary influence on our doxastic attitudes, and probably over our acts of imagination. We can train ourselves to be more critical of things like gossip. We might also train ourselves to be

[2] By 'responsible' we mean that one is the proper subject of praise, blame, or neutral appraisal.

more imaginative: perhaps by engaging with imaginative people and their work, or simply by trying to conceive of imaginative solutions to everyday problems. According to William Alston, because deontologism needs to be "grounded" on our indirect control of indirect influence, it has to be a thesis about blamelessness. For, as Weatherson himself notes, in such a formulation what we are responsible for are not particular doxastic tokens, but rather certain actions such as training oneself to be less credulous. But of course, as Alston continues and Weatherson does not, this does not mean that we cannot be held to *blame* for particular doxastic tokens. This is because blame supervenes on requirement in two ways, as Alston puts it:

> First, and most simply, one is to blame for failing to do something required. But second, one is to blame for the obtaining of some fact if that fact would not have obtained if one had not behaved in some manner for which one is to blame in the first sense, that is, for doing something forbidden or failing to do something required.[3]

Of course, the same can be said for praise. Thus, this point counts in favor of neither DDP nor DDB. What is important, however, is that we can talk of praise or blame for particular doxastic tokens without needing to "outrun" voluntary control, since we can base such judgments on our indirect doxastic control or indirect doxastic influence. So our first worry with this argument in favor of DDP is that it provides a solution to a problem that already has an equally plausible solution.

In bringing the notion of indirect voluntary control into play, we see Weatherson's example in quite a different light. We can now wonder whether the praiseworthiness we ascribe to the cricket captain in fact is due to his ability to engage and train his imagination. It might also be that the endowment of such praise partly depends on whether the act of imagination yields a particular result or arrives at a propitious moment; would we still praise the captain if his field placement involved a strong element of risk in a close, tense game? The ability to engage one's imagination at appropriate times may well be something over which we have indirect control. Our attribution of praise, when it comes to the imagination, may be determined by a variety of factors, many of which involve indirect voluntary control.

Second, there seem to be two different kinds of praise. On the one hand, we can praise some person S for φ-ing if we value S's φ-ing, without holding S responsible for φ-ing. Thus, I might praise Miranda for her beauty without holding her responsible for it, and I may praise

[3] William P. Alston, *Epistemic Justification: Essays in the Theory of Knowledge* (Ithaca: Cornell, 1989), p. 140.

my recently bought Chevrolet for its speed. This sort of praise ought to be distinguished clearly from that usually classified as what Peter Strawson calls a "participant reactive attitude,"[4] entailing that we hold the person in question responsible. In this sense, we can *in principle* also, say, blame or resent *S* for φ-ing.

Now, if the imaginative cricket captain did not exercise his capacity to train his imagination but happened to come up with imaginative field placements nonetheless, we would praise him only in the first sense of the word. But this sense has nothing to do with responsibility: we just value the captain's imaginative field placements and admire him because of that. But this means that this example is irrelevant here.

Third and finally, if, contrary to what we have suggested, the captain *is* praiseworthy in the second sense, why would we think that all responsible believing is like the captain's imaginative act? Surely, if praiseworthy beliefs are of that sort, they provide a poor model for deontological doxastic justification. On doxastic justification, so understood, the standards of justification are too high, since very few of our beliefs are justified.

II. DEONTOLOGICAL INTUITIONS AND DEMONIC DECEPTION

In his article, Weatherson discusses the so-called New Evil Demon (NED) problem, not in order to solve it, but as a means to show that the deontological intuitions that play a role in NED scenarios favor DDP over DDB. The NED problem usually is presented as a problem for reliabilist theories of doxastic justification. The thought is that I am equally justified in my beliefs as is my doxastic counterpart (who has exactly the same beliefs, memories, and intuitions I have and is disposed to reason as I do) even though, unlike me, he lives in a world governed by an evil demon who systematically deceives him, and so happens to have (at least mostly) false beliefs. Consider James Pryor's way of thinking about the problem. Assume that there are three victims of equal demonic deception *A*, *B*, and *C*. Victim *A* frequently uses faulty reasoning procedures to arrive at her beliefs, and if she were a little more careful she would easily see that they are faulty. Victim *B* also often uses faulty procedures to arrive at her beliefs, but the faults in the procedures are the product of a bad upbringing, and the mistakes are so subtle that we cannot reasonably expect *B* to notice them. Victim *C* hardly ever uses faulty procedures to arrive at her beliefs; in fact, she displays the paradigm of

[4] Peter F. Strawson, "Freedom and Resentment," in *Freedom and Resentment and Other Essays* (New York: Harper & Row: 1974), pp. 1–25.

good reasoning but still has mostly false beliefs because she is the victim of devilish deception.[5]

Following Pryor, Weatherson claims that the notion of blameworthiness cannot capture the intuition that *C* is epistemically better than *B*. If both are blameless, then both are equally justified according to DDB. Weatherson's solution to Pryor's version of the NED problem is to claim that none of the victims are justified, but that that does not preclude us from attributing praise, and in this way we can accommodate all the intuitions at play in the scenario. Let us explain. According to Weatherson, *A*, *B*, and *C* are all bad *evidence collectors*: their evidence is misleading evidence. Therefore, the beliefs of these three victims are unjustified. However, there is an important distinction to be made between *A* and *B* on the one hand and *C* on the other, namely, that only *C* is a good *evidence processor*: she processes her evidential input excellently. Thus, *C* is epistemically praiseworthy in a way that *A* and *B* are not. Thus, according to Weatherson, the attribution of praise helps us make sense of the difference between *B* and *C*, since we can say that *C* is epistemically praiseworthy (though unjustified) in having the beliefs that she does in a way that *B* is not.[6]

In response to this argument, we first should notice that Weatherson departs from his original understanding of doxastic justification in terms of praiseworthiness, as expressed in DDP (in conjunction with DD). It now seems that one is justified only if one's evidence is not misleading, and whether one's evidence is misleading need not be up to oneself. On this alternative understanding of DDP, justification cannot be understood in deontological terms only. This means that, strictly speaking, we have to revise DDP and DDB as follows:

(DDP*) *S* responsibly (justifiedly) believes that *p only if S* is praiseworthy for believing that *p*,

and, presumably,

(DDB*) *S* responsibly (justifiedly) believes that *p only if S* is blameless for believing that *p*.

However, since all of our criticisms in this paper are directed against the view that praiseworthiness for believing that *p* is a *necessary* condition for responsibly believing that *p*, we can continue to work with DDP and DDB.

[5] Cf. James Pryor, "Highlights of Recent Epistemology," *The British Journal for the Philosophy of Science*, LII, 1 (2001): 95–124, at p. 117.

[6] Similarly, Weatherson argues that a person who displays the virtue of cosmopolitanism is praiseworthy in a way that someone who, albeit blamelessly, displays the vice of patriotism is not (p. 566).

Now, what about the difference between victims *A* and *B*? It seems impossible to articulate the difference between them if we only have the concept of praiseworthiness at hand, since neither is worthy of praise. Yet there is a difference between them, namely, that *B* is blameless while *A* is not. So in order to make sense of all the intuitions in this scenario, deontology must not merely be about praise; it has to be about blame, too. And now the question is how we determine when someone is a "good" doxastic agent, "good" such that it is necessary for doxastic justification. We can either draw the boundary between what counts as a good doxastic agent and what does not, high or low (that is, we can say that only a praiseworthy doxastic agent is a good one), or we can say that a doxastic agent need only be blameless to qualify as "good." Either way, we will not be able to accommodate all the intuitions in the skeptical scenario above. So the ability to meet that constraint cannot help us decide between DDP and DDB. Since whichever way we take deontologism leaves an intuition unaccounted for, this is also problematic for Weatherson's overall solution to the NED problem. For what advantage is left to his solution?[7]

III. A CONTRADICTION FROM DDB?

Let us now turn to a third and final consideration in favor of DDP, as it is formulated by Weatherson. Here, the idea is that the conjunction of DDB (in contradistinction from DDP), DD, and the plausible premises

> A_1: It is possible for *S* to have a justified but false belief that her belief that *p* is justified. (567)
> A_2: If *S* blamelessly believes that she is justified in believing that *p*, and on the basis of that belief comes to believe that *p*,[8] then she is blameless in believing that *p*. (568)

leads to the following contradiction:

(1) *S* justifiedly, but falsely, believes that she is justified in believing that *p*. (A_1)
(2) On the basis of this belief, *S* comes to believe that *p*. (Assumption)

[7] Recall that Weatherson's motivation for it is that "[a] fairly common response is to note that even according to externalist epistemology there will be some favorable property that the victim's beliefs have, and this can explain the intuition that there is something epistemically praiseworthy about the victim's beliefs. My approach is a version of this, one that is invulnerable to recent criticisms of the move" (p. 564).

[8] We are not quite sure how to understand this. Does Weatherson mean that *S* comes to believe that *p* on the basis of thinking (a) that *a* belief that *p* is justified, (b) that her belief that *p would be* justified if she were to acquire it, or (c) that the belief that she already has *is* justified? It seems highly doubtful that one can come to believe something one already believes, so, for the sake of charity, we take Weatherson to have in mind either (a) or (b). We do not think anything hangs on which of the two one takes.

(3) *S* blamelessly believes that she is justified in believing that *p*. (1, J=B [DDB])
(4) *S* blamelessly believes that *p*. (2, 3, A_2)
(5) *S* is justified in believing that *p*. (4, J=B [DDB])
(6) It is false that *S* is justified in believing that *p*. (1) (Cf. 568–69)

Given that (5) and (6) are logically contradictory, so the argument goes, we have to reject DDB (Weatherson's J=B), A_1, or A_2. A_2, says Weatherson, "is extremely plausible," so either A_1 or DDB has to go. But to give up A_1 is to commit oneself to externalism, so if we want to be deontologists and internalists we had better give up DDB in favor of DDP.

Before responding to this argument against DDB, let us first consider whether DDP fares any better. On Weatherson's proposal, this means, first, that we replace DDB with DDP, and second, that we say that the inference from *I am justified in believing that p* to *p* is itself praiseworthy only if the premise (that is, that *I am justified in believing that p*) is true. It is not entirely clear what this is supposed to mean (does it mean to say something merely about the epistemic status of the inference to *p* or the epistemic status of the belief that *p* itself?) The idea, however, seems to be that we should replace A_2 with A_3:

> A_3: If *S* is praiseworthy for believing that she is justified in believing that *p*, and on the basis of that belief comes to believe that *p*, then she is praiseworthy for believing that *p* only if *S* is justified in believing that *p*.[9]

Now, what does the "argument" on Weatherson's alternative look like? He fails to lay it out, but it seems that it would look as follows:

(7) *S* justifiedly, but falsely, believes that she is justified in believing that *p*. (A_1)
(8) On the basis of this belief, *S* comes to believe that *p*. (Assumption)
(9) *S* is praiseworthy for believing that she is justified in believing that *p*. (7, J=P [DDP])
(10) *S* is *not* praiseworthy for believing that *p*. (8, 9, A_3)
(11) *S* is *not* justified in believing that *p*. (10, J=P [DDP])
(12) It is false that *S* is justified in believing that *p*. (7)

And, clearly, there is no contradiction involved in this set of propositions. The problem with A_3, however, is that it seems trivially true. Given DDP (Weatherson's J=P), A_3 could be rephrased as

> A_{3*}: If *S* is praiseworthy for believing that she is praiseworthy for believing that *p*, and on the basis of that belief comes to believe that *p*, then

[9] In Weatherson's own words: "First, we say that a belief's being justified is not a matter of it being blameless, but a matter of it being in a certain way praiseworthy. Second, we say that the inference from *I am justified in believing that p* to *p* is not praiseworthy if the premise is false" (p. 569).

she is praiseworthy for believing that *p* only if *S* is praiseworthy for believing that *p*.

A_3 is true by definition, in the same way as A_{2*} is true by definition:

A_{2*}: If *S* blamelessly believes that she is blameless for believing that *p*, and on the basis of that belief comes to believe that *p*, then she is blameless for believing that *p* only if she is blameless for believing that *p*.

The really important question that Weatherson should have addressed is whether the following genuinely informative principle is true:

A_4: If *S* is praiseworthy for believing that she is justified in believing that *p*, and on the basis of that belief comes to believe that *p*, then she is praiseworthy for believing that *p*,

or

A_{4*}: If *S* is praiseworthy for believing that she is praiseworthy for believing that *p*, and on the basis of that belief comes to believe that *p*, then she is praiseworthy for believing that *p*.

And the problem is that A_4/A_{4*} seems equally plausible as A_2/A_{2*}. If praiseworthiness is transferred from a meta-belief *B** to a belief *B* in virtue of *B**'s being about *B* in a specific way, then why would blamelessness not be transferred from a meta-belief *B** to a belief *B* in virtue of *B**'s being similarly about *B*? Hence, Weatherson's reasoning at this point fails to favor DDP over DDB.

Second, this argument against DDB fails quite simply on the falsity of Weatherson's assumption that we need to be externalists in order to think that we have infallible knowledge about our beliefs about justification, since we could just stipulate an internalist condition on top of the infallibility condition. At most, the necessity of having infallible beliefs about justification implies a strong form of access internalism according to which only reasons that I have special, infallible access to (through introspection) can justify for me a belief that *p*. Further, one need only be committed to the claim that not all justified beliefs are true, or that justification does not *entail* truth, and this is quite consistent with thinking that justification entails truth when it comes to propositions about one's own reasons. So we have not eschewed the internalist idea that one can have justified but false beliefs. Weatherson seems to anticipate this objection:

Now some may think that the general principle [A_1; authors] is right, but that beliefs about what we are justified in believing are special, and if they are justified they are true. But such an exception seems intolerably ad hoc. If we can have false but justified beliefs about some things, then presumably we can have false but justified beliefs about our evidence,

> since in principle our evidence could be practically anything. So the following situation seems possible; indeed it seems likely that something of this form happens frequently in real life. *S* has a false but justified belief that *e* is part of her evidence. *S* knows both that anyone with evidence *e* is justified in believing *p* in the absence of defeaters, and that there are no defeaters present. So *S* comes to believe, quite reasonably, that she is justified in believing that *p*. But *S* does not have this evidence, and in fact all of her evidence points toward ~*p*. So it is false that she is justified in believing *p*. (567–68)

Here, Weatherson claims that there are situations in which (a) some person *S* has a false but justified belief that *e* is part of her evidence; (b) *S* knows that anyone with evidence *e* is justified in believing *p* in the absence of defeaters; (c) *S* knows that she has no defeaters for *e* (or for believing that *p*); (d) *S* thereby comes to believe that she is justified in believing that *p*; (e) *S* lacks *e*; (f) *S*'s evidence strongly points toward not-*p*; and, therefore, (g) *S* is not justified in believing *p*. Weatherson's argument, however, trades on the ambiguity of the word 'justification', which he uses purely deontologically and internalistically in (a), (b), and (d), but externalistically in (g). It is because of this ambiguity that we can imagine a situation along these lines. If, however, we understand all instances of 'justified' in this example purely internalistically and deontologically—as we ought to, if DDB is the thesis under investigation—we see that no such scenario is possible. It seems impossible that some person *S* (i) has a blameless belief that *e* is part of her evidence, (ii) knows that anyone with evidence *e* is blameless in believing that *p* in the absence of defeaters and that she has no defeaters for *e* (or for believing that *p*), (iii) on the basis of that comes to believe that *p*, and (iv) is blameworthy for believing that *p*. If conditions (i)–(iii), then for *S* to believe responsibly is clearly to believe that *p*, whether or not her believing that *p* is justified in any externalist sense of the word.

IV. CONCLUSION

For all Weatherson shows, there is no reason to prefer DDP to DDB. All three arguments against DDB fail at one or several junctures. Moreover, as we have seen in response to the voluntarism problem and the New Evil Demon problem, we should not set the standards of justification too high if we do not want to risk skepticism. DDB seems much more capable of meeting this demand than DDP. This provides us with sufficient reason to adhere to the standard account of responsible belief in terms of blamelessness rather than praiseworthiness.

ANTHONY BOOTH
RIK PEELS

Utrecht University

BOOK REVIEWS

Real Materialism and Other Essays. GALEN STRAWSON. New York: Oxford University Press, 2008. x + 478 p. Cloth $130.00, paper $45.00.

Galen Strawson is an exceptionally good philosopher. He is a clear and sometimes stylish writer; he is patient and meticulous in his presentation of arguments; he is independent-minded, unafraid to buck the trend; he tackles important issues; he has a fine historical sense; he has excellent judgment; and he is right an awful lot of the time (especially when he is bucking the trend). The nineteen essays collected in this volume display all these qualities vividly, and together they constitute a solid and enduring body of work. The topics covered range widely, falling into six (overlapping) groups: physicalism and experience, color, the self, intentionality, free will and moral responsibility, and Hume on causation. I will discuss Strawson's treatment of these topics in reverse order, beginning with the essays I agree with most and ending with those where I have some disagreements.

The three essays on Hume extend Strawson's 1989 book *The Secret Connexion*, which argued for a novel interpretation of Hume. By now the exegetical line adopted by Strawson (and by Edward Craig and John Wright) has become familiar, but earlier it was quite heterodox. Before I read the book for review,[1] I accepted the orthodox interpretation—that Hume wholly rejects casual necessity "in the objects," holding instead that there is nothing in reality except regular succession—but studying Strawson's book converted me to the "skeptical realist" interpretation of Hume, according to which Hume's point is entirely epistemological. There is indeed nothing more to our *knowledge* of causation than regular succession, since we have no impression of necessary connection, but Hume does not doubt that there really is necessity in the objects. The key is to distinguish between what we find intelligible in the light of our impressions and what we can merely refer to without grasping its essential nature—between what Berkeley called Ideas and Notions, roughly. Objects doubtless have powers, as we naturally believe, but it is an error to think that we have any positive descriptive understanding of what

[1] The review appeared as "Reputation" in the *London Review of Books*, XI, 22 (November 23, 1989): 6–7.

0022-362X/10/0705/266–271

these powers consist in; we have a referential *notion* but no substantive *idea* of power. Thus Hume is a realist about causation but a skeptic about our insight into its nature. In much the same way, he does not dispute that external objects exist, but he doubts that we know much about their nature—they are, for us, just whatever cause our perceptions.

Strawson's defense of this interpretation of Hume is scrupulous, sensitive, comprehensive, and utterly convincing. What I am not clear about, though, is whether he agrees with Hume, though he is obviously sympathetic: does he think that our concept of causation extends no further than regular succession, and if so does he accept Hume's empiricist account of concepts? I doubt the latter, but then it is a question how to sustain Hume's skeptical view without it. I think myself that we do have a satisfactory positive conception of objective causal necessity—though I would agree that the springs of such necessity in the underlying forces of nature (such as gravity) are opaque to us. I would like to see Strawson write an article telling us what his own view of causation is in the light of this interpretation of Hume.

On free will Strawson defends a straightforwardly ontological doctrine: there is no such thing. Here he relies on the *causa sui* argument, to the effect that genuine freedom—the kind that would ground moral responsibility—requires the ability to choose one's nature, but this is impossible given that any such choice presupposes a prior nature. This is a powerful argument, saddling the idea of freedom with an impossible ontological obligation—meaningful choice independent of anything that could confer meaning upon it. The tabula rasa cannot choose its own moral character—for what could such a choice issue from? Strawson develops the argument with characteristic care and resourcefulness, and he points out that the dispute between determinists and indeterminists is quite irrelevant to the heart of the issue. Whether determinism is true or not, agents cannot be their own cause and so cannot be held responsible for what their nature leads them to do. You might have thought that the free will debate had run out of steam by now, every option having been exhaustively investigated, but Strawson breathes fresh life into it—if not into free will itself. My only reservation about his discussion is that he gives so little attention to the analogue of Hume's view of causation: there doubtless is free will but we are unable to frame an adequate idea of what it inherently involves.[2]

[2] I discuss this possibility in *Problems in Philosophy: The Limits of Inquiry* (Malden, MA: Blackwell: 1993), chapter 5.

The self also comes in for a bit of a battering in the four essays devoted to discovering what it is and how long it lasts. Or better: the enduring self is treated with suspicion, but the momentary self is awarded a central metaphysical position. On the one hand, every experience presupposes a self that the experience is *for*, so that the self is no construct from more primitive psychological elements; yet, on the other hand, it can scarcely survive a night's sleep, lasting "for a maximum of about two seconds" (161). (This is what Strawson calls the "thin subject," as opposed to the "thick subject," who is just the human being that houses the succession of fleeting selves.) He also contrasts a diachronic and episodic understanding of the self, where this is a matter, not of what the persistence of the self actually is, but of how we tend to experience our lives. Some people, he reports, feel strongly that they were around in the far past and will be in the distant future, while others (such as Strawson himself) feel very little sense of continuity with the self that occupied temporally remote parts of their life. He also discusses the question of whether people regard their lives as a *narrative* of some sort, again pouring cold water on the idea. These are interesting discussions, which again introduce new ideas into the question of personal identity. My own feeling is that I am not sure I understand the question: if you ask me whether I experience myself as the same self that was around in my vicinity forty or fifty years ago, I just do not know what to say, except "well, in some ways they are similar but in others different." The dispute between the diachronic and the episodic self-conceptions seems underdetermined by any fact that might be appealed to, rather like the question of whether the England of the sixteenth century is the same as the England of the twentieth century; "yes and no" seems the only sensible answer. I do, however, agree with Strawson that the notion of a psychological subject is unavoidable in some form, because written into the very structure of experience; but what it is, metaphysically, is fiercely hard to grasp—and here again a Humean skeptical realist approach can seem attractive.

Intentionality also emerges as fully real yet closely confined: experiences possess intentionality as robustly as, say, birds possess wings, but *nothing else does*—including beliefs! Only states of consciousness (which includes occurrent thoughts) have genuine intentionality, for Strawson, but beliefs are dispositions—the kind of thing you can have while unconsciously asleep—and so they are not really about anything. Strawson has two main arguments for this striking thesis: (a) beliefs are dispositions and dispositions are not the metaphysically right kind of thing to have intentionality (just as a disposition to fragility is not the kind of thing that can be broken), and (b) unless we

hold the line at states of consciousness we run the risk of spreading intentionality too widely. Once we allow that nonconscious states can have intentionality we end up letting thermometers and the like in too, with the attendant threat that all intentionality is merely as-if. It is either Dennett or experiential confinement, with no middle ground. This is pretty shattering stuff and not the view I have accepted lo these many years—yet Strawson puts up a surprisingly strong case for it. Of course, we can speak *as if* beliefs have intentionality, since they are dispositions that give rise to states of consciousness that indisputably have it, namely, conscious thoughts; still, nothing has nonderivative intentionality except experiences (in the wide sense Strawson favors, which includes cognitive phenomena). This greatly simplifies the picture, to be sure, and lets us off many uncomfortable attributions of (nonderivative) intentionality, for example, to brain states. I commend this discussion to all readers with an interest in discovering where real aboutness begins and ends; I might even become convinced of it myself in due course.

On one question, though, I find myself unmoved by Strawson's bracing iconoclasm, and this is with respect to a topic on which he seems to me, ironically, to cleave too closely to orthodoxy. In "Red and 'Red'" he comes to the conclusion that "red" does not denote a color or any other property. This is not because he does not think predicates denote properties, but rather because of specific features of color terms. To simplify a (suspiciously) complex argument, he holds that color terms have their meaning fixed by their public use, a la Wittgenstein, but that the qualitative character of a speaker's experience is too subjective and variable to sustain a single property as what might be denoted by the term. In a color inversion case, say, the use is the same, though the experience varies; so the meaning has to be identical, despite the subjective variation. It seems to me that Strawson here fails to consider the most plausible option for describing such a case, namely, that "red" denotes the disposition objects have to produce the kind of experience the speaker has when looking at certain objects (or some primitive color property supervenient on such a disposition).[3] This does indeed make the meaning of the term vary with the qualitative character of the speaker's experience, and introduces a deep privacy into meaning; but I do not see why we should be swayed against this by Wittgensteinian insistence that all meaning (and all *of* meaning) must be public, any more than

[3] I discuss this kind of theory in "Another Look at Color," this JOURNAL, XCIII, 11 (November 1996): 537–53.

experience itself must be (which Strawson concedes to be private in the relevant sense). There is really nothing wrong, in the end, with the view that you and I might mean something different by "red," despite our agreement in its public use, simply because we denote different dispositions to cause experience by the term (couldn't I just *stipulate* that I am using "red" to stand for the disposition to cause precisely the experience I am having now?). This is the one instance in which I think Strawson goes awry because of allegiance to a questionable orthodoxy.

The volume begins with the long paper "Real Materialism," followed by "Realistic Monism: Why Physicalism Entails Panpsychism," in which Strawson defends the radical irreducibility of experience, a version of panpsychism, the limits of physics as a theory of the (physical) world, *and materialism*. Say what? You heard correctly: materialism/physicalism is true, he insists, even though physics (and physiology, and so on) do not and cannot specify the nature of experience and even fall short of giving the nature of matter, electricity, gravitation, and so on—where this latter nature is best conceived as a kind of "mind stuff." That does not sound much like the materialism you learned about in graduate school, what with all the irreducible experiences and the panpsychist world and the incompleteness of physics as an account of concrete reality at the most basic level (atoms and whatnot). Strawson is well aware of this but remains unbowed: his is a "materialism" of a very idiosyncratic stripe, a terminological decision that is held to be metaphysically important. He makes many good points along the way about why experience is irreducible, why physics is a partial account of physical reality, and even why panpsychism has a lot going for it. Still, I fail to see why he insists on describing his position as "materialist," given that he has bleached the term of all descriptive content. He firmly dissociates his materialism from the thesis that the physical sciences describe all of (concrete) reality, but then it seems left with no content at all. The word seems to function purely honorifically in his mouth. Given his general position, I think he would do better to drop the label "materialist" altogether, since it no longer has any clear meaning (as Chomsky has argued)—an option he contemplates but rejects. Nor do I think he should call himself a monist, since he makes a firm distinction between experiential and nonexperiential reality. His argument for panpsychism will then be simply that the irreducibility of experience requires that reality be experiential all the way down, instead of the odd-sounding claim that physicalism entails panpsychism. (On the normal way of understanding these terms, of course, these two doctrines are sharply opposed.) Still, this terminological issue aside, these

papers are stimulating and often convincing; I particularly like his discussion of the Eddington-Russell interpretation of physics as providing merely "structural" information about reality, with the real nature of physical reality left unspecified, which poses a serious threat to materialist metaphysics.[4]

COLIN MCGINN

University of Miami

[4] I have a much fuller discussion of Strawson's views on physicalism and panpsychism in "Hard Questions," my reply to him in *Consciousness and Its Place in Nature*, ed. Anthony Freeman (Charlottesville, VA: Imprint Academic, 2006), pp. 90–99.

NEW BOOKS

BARTHA, PAUL: *By Parallel Reasoning.* New York: Oxford, 2010. x, 356 p. Cloth $74.00.

BEERE, JONATHAN: *Doing and Being: An Interpretation of Aristotle's Metaphysics Theta.* New York: Oxford, 2010. xii, 367 p. Cloth $86.00.

BEISER, FREDERICK C.: *Diotima's Children: German Aesthetic Rationalism from Leibniz to Lessing.* New York: Oxford, 2010. x, 295 p. Cloth $80.00.

BELSHAW, CHRISTOPHER: *Annihilation: The Sense and Significance of Death.* Montreal: McGill-Queen's, 2010. xii, 258 p. Cloth $95.00, Paper $27.95.

BHASKAR, ROY: *Plato Etc.: The Problems of Philosophy and their Resolution.* New York: Routledge, 2010. xxx, 267 p. Cloth $115.00, Paper $47.95.

BORTOLOTTI, LISA: *Delusions and Other Irrational Beliefs.* New York: Oxford, 2010. 298 p. Cloth $65.00.

BUCHANAN, ALLEN: *Human Rights, Legitimacy, & the Use of Force.* New York: Oxford, 2010. viii, 332 p. Cloth $74.00.

CARLSON, LICIA: *The Faces of Intellectual Disability: Philosophical Reflections.* Bloomington: Indiana UP, 2010. xiii, 266 p. Cloth $65.00, Paper $24.95.

CURRIE, GREGORY: *Narratives & Narrators: A Philosophy of Stories.* New York: Oxford, 2010. 243 p. Cloth $55.00.

D'AGOSTINO, FRED: *Naturalizing Epistemology: Thomas Kuhn and the 'Essential Tension.'* New York: Palgrave MacMillan, 2010. viii, 216 p. Cloth $85.00.

FRANKS, CURTIS: *The Autonomy of Mathematical Knowledge: Hilbert's Program Revisited.* New York: Cambridge, 2009. xiii, 213 p. Cloth $75.00.

GALE, RICHARD M. GALE: *John Dewey's Quest for Unity: The Journey of a Promethean Mystic.* Amherst, NY: Prometheus Books, 2010. 215 p. Cloth $32.98.

GAUT, BERYS: *A Philosophy of Cinematic Art.* New York: Cambridge, 2010. xiii, 324 p. Cloth $99.00, Paper $29.99.

GEUSS, RAYMOND: *Politics and the Imagination.* Princeton: University Press, 2010. xvi, 198 p. Paper $24.95.

GLYMOUR, CLARK: *Galileo in Pittsburgh.* Cambridge: Harvard, 2010. x, 150 p. Cloth $27.95.

GOTZ, IGNACIO L.: *Conceptions of Happiness.* Lanham, MD: University Press of America, 2010. xxiv, 223 p. Paper $35.00.

GROSS, MICHAEL L.: *Moral Dilemmas of Modern War: Torture, Assassination, and Blackmail in an Age of Asymmetric Conflict.* New York: Cambridge, 2010. xiii, 321 p. Cloth $85.00, Paper $27.99.

HELM, BENNETT W.: *Love, Friendship, and the Self: Intimacy, Identification, and the Social Nature of Persons.* New York: Oxford, 2010. xiii, 316 p. Cloth $75.00.

HILL, LESLIE: *Radical Indecision: Barthes, Blanchot, Derrida, and the Future of Criticism.* Notre Dame: University Press, 2010. xiv, 438 p. Paper $45.00.

KAZEZ, JEAN: *Animalkind: What We Owe to Animals.* Malden, MA: Wiley-Blackwell, 2010. viii, 206 p. Cloth $89.95, Paper $24.95.

KRISTJANSSON, KRISTJAN: *The Self and Its Emotions.* New York: Cambridge, 2010. xiv, 272 p. Cloth $85.00.

LEBUFFE, MICHAEL: *From Bondage to Freedom: Spinoza on Human Excellence.* New York: Oxford, 2010. viii, 253 p. Cloth $74.00.

LONERGAN, ERIC: *Money*. Durham, UK: Acumen, 2010. vii, 149 p. Paper $18.95.

LOPTSON, PETER: *Reality: Fundamentals Topics in Metaphysics*. Ottawa, Ontario: Ottawa, 2010. viii, 346 p. Paper $39.00.

MILLER, SEUMAS: *The Moral Foundations of Social Institutions: A Philosophical Study*. New York: Cambridge, 2010. x, 371 p. Cloth $90.00, Paper $28.99.

MOSER, PAUL K.: *The Evidence for God: Religious Knowledge Reexamined*. New York: Cambridge, 2010. x, 280 p. Cloth $85.00, Paper $24.99.

NARVESON, JAN: *This Is the Ethical Theory*. Chicago: Open Court, 2010. x, 283 p. Paper, $36.95.

PEIRCE, CHARLES S.: *Writings of Charles S. Peirce: A Chronological Edition, Volume 8 1890-1892*. Bloomington: Indiana UP, 2010. xcvii, 724 p. Cloth $59.95.

REINERS, WLLIAM A. AND JEFFREY A. LOCKWOOD: *Philosophical Foundations for the Practices of Ecology*. New York: Cambridge, 2010. xiii, 212 p. Cloth $110.00, Paper $45.00.

SEARLE, JOHN R.: *Making the Social World: The Structure of Human Civilization*. New York: Oxford, 2010. xiv, 208 p. Cloth $24.95.

SEUREN, PIETER A. M.: *The Logic of Language*. New York: Oxford, 2010. xv, 428 p. Cloth $99.00.

SLOTE, MICHAEL: *Essays on the History of Ethics*. New York: Oxford, 2010. 165 p. Paper $65.00.

SMITH, BARBARA HERRNSTEIN: *Natural Reflections: Human Cognition at the Nexus of Science and Religion*. New Haven: Yale, 2010. xviii, 201 p. Cloth $28.00.

STALEY, GREGORY A.: *Seneca and the Idea of Tragedy*. New York: Oxford, 2010. xiii, 185 p. Paper $74.00.

STEARS, MARC: *Demanding Democracy: American Radicals in Search of a New Politics*. Princeton: University Press, 2010. x, 246 p. Cloth $29.95.

STOKES, PATRICK: *Kierkegaard's Mirrors: Interest, Self, and Moral Vision*. New York: Palgrave MacMillan, 2010. xii, 223 p. Cloth $85.00.

SWEETMAN, BRENDAN: *Religion and Science: An Introduction*. New York: Continuum, 2010. viii, 170 p. Cloth $24.95.

ULEMAN, JENNIFER K.: *An Introduction to Kant's Moral Philosophy*. New York: Cambridge, 2010. 189 p. Cloth $85.00, Paper $28.99.

VOORHOEVE, ALEX: *Conversations on Ethics*. New York: Oxford, 2010. 259 p. Cloth $34.95.

WEIRICH, PAUL: *Collective Rationality: Equilibrium in Cooperative Games*. New York: New York, 2010. xi, 270 p. Cloth $74.00.

WINFIELD, RICHARD DIEN: *Hegel and Mind: Rethinking Philosophical Psychology*. New York: Palgrave MacMillan, 2010. xiii, 170 p. Cloth $85.00.

YAGISAWA, TAKASHI: *Worlds and Individuals, Possible and Otherwise*. New York: Oxford, 2010. 314 p. Cloth $85.00.

ZAGORIN, PEREZ: *Hobbes and the Law of Nature*. Princeton: University Press, 2009. x, 177 p. Cloth $29.95.

ZARETSKY, ROBERT and JOHN T. SCOTT: *The Philosopher's Quarrel: Rousseau, Hume, and the Limits of Understanding*. New Haven: Yale, 2010. x, 247 p. Paper $18.00.

crítica

REVISTA HISPANOAMERICANA DE FILOSOFÍA

Vol. 42 / No. 124/ abril 2010 / ISSN 0011–1503

Crítica, Revista Hispanoamericana de Filosofía aparece tres veces al año en los meses de abril, agosto y diciembre. Es una publicación del Instituto de Investigaciones Filosóficas (UNAM). Todo tipo de correspondencia deberá ser dirigida al Apartado Postal 70-447, Coyoacán, 04510, México D.F. Correo electrónico: critica@filosoficas.unam.mx
Versión electrónica: http://critica.filosoficas.unam.mx

UNITED STATES POSTAL SERVICE®

Statement of Ownership, Management, and Circulation (All Periodicals Publications Except Requester Publications)

1. Publication Title	2. Publication Number	3. Filing Date
The Journal of Philosophy	0 0 2 2- 3 6 2 X	9/26/2010
4. Issue Frequency	**5. Number of Issues Published Annually**	**6. Annual Subscription Price**
Monthly	12	Individuals: $45.00 ~~Institutions: $100.00~~

7. Complete Mailing Address of Known Office of Publication *(Not printer) (Street, city, county, state, and ZIP+4®)*
The Journal of Philosophy, Columbia Univ.
1150 Amsterdam Ave, Rm 709
New York, NY 10027

Contact Person: Pam Ward
Telephone *(Include area code)*: 212.666.4419

8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not printer)*
Same

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*

Publisher *(Name and complete mailing address)*
The Journal of Philosophy, Inc.
1150 Amsterdam Ave
New York, NY 10027

Editor *(Name and complete mailing address)*
Akeel Bilgrami, President
The Journal of Philosophy
1150 Amsterdam Ave New York, NY 10027

Managing Editor *(Name and complete mailing address)*
Alyssa Timin, Managing Editor
The Journal of Philosophy 1150 Amsterdam Ave
New York, NY 10027

10. Owner *(Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)*

Full Name	Complete Mailing Address
The Journal of Philosophy	Columbia University 1150 Amsterdam Ave Mc 4972 New York, NY 10027

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☐ None

Full Name	Complete Mailing Address

12. Tax Status *(For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)*
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☐ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months *(Publisher must submit explanation of change with this statement)*

PS Form **3526,** September 2007 *(Page 1 of 3 (Instructions Page 3))* PSN 7530-01-000-9931 **PRIVACY NOTICE:** See our privacy policy on *www.usps.com*

13. Publication Title	14. Issue Date for Circulation Data Below
The Journal of Philosophy	May 2010

15. Extent and Nature of Circulation			Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies *(Net press run)*			3500	3500
b. Paid Circulation *(By Mail and Outside the Mail)*	(1)	Mailed Outside-County Paid Subscriptions Stated on PS Form 3541(Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)	700	674
	(2)	Mailed In-County Paid Subscriptions Stated on PS Form 3541 *(Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)*	2400	2247
	(3)	Paid Distribution Outside the Mails Including Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Paid Distribution Outside USPS®		
	(4)	Paid Distribution by Other Classes of Mail Through the USPS (e.g. First-Class Mail®)	3100	2921
c. Total Paid Distribution *(Sum of 15b (1), (2), (3), and (4))*				
d. Free or Nominal Rate Distribution *(By Mail and Outside the Mail)*	(1)	Free or Nominal Rate Outside-County Copies included on PS Form 3541		
	(2)	Free or Nominal Rate In-County Copies Included on PS Form 3541		
	(3)	Free or Nominal Rate Copies Mailed at Other Classes Through the USPS (e.g. First-Class Mail)		
	(4)	Free or Nominal Rate Distribution Outside the Mail *(Carriers or other means)*		
e. Total Free or Nominal Rate Distribution *(Sum of 15d (1), (2), (3) and (4))*			150	150
f. Total Distribution *(Sum of 15c and 15e)*			3250	3071
g. Copies not Distributed *(See Instructions to Publishers #4 (page #3))*			250	429
h. Total *(Sum of 15f and g)*			3500	3500
i. Percent Paid *(15c divided by 15f times 100)*			95.45	95.12

16. Publication of Statement of Ownership
☐ If the publication is a general publication, publication of this statement is required. Will be printed in the 05/2010 issue of this publication. ☐ Publication not required.

17. Signature and Title of Editor, Publisher, Business Manager, or Owner
Pamela C. Ward — Date: 9/20/2010

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

PS Form **3526,** September 2007 *(Page 2 of 3)*