Philosophy and Phenomenological Research
Vol. LXIII, No. 1, July 2001

Mental Causation: Unnaturalized but not Unnatural

ERIC MARCUS
Auburn University

The central problem for a realist about mental causation is to show that mental causation is compatible with the causal completeness of physical systems. This problem has seemed intractable in large part because of a widely held view that any sort of systematic overdetermination of events by their causes is unacceptable. I account for the popularity of this view, but argue that we ought to reject it. In doing so, I show how we thereby undermine the idea that mental causes *must* be naturalizable in order to be legitimate. Thus I argue that a non-naturalist conception of mental causation is compatible with a plausible kind of physicalism.

If a woman in the audience at a presentation raises her hand, we would take this as evidence that she intends to ask a question. In normal circumstances, we would be right to say that she raises her hand *because* she intends to ask a question. We also expect that there could, in principle, be a causal explanation of her hand's rising in purely physiological terms. Ordinarily, we take the existence and compatibility of both kinds of causes for granted. But this can come to seem strange. When we imagine tracking the physiological process that culminates in her hand's rising, it is hard to find a purchase for her intention. The physiological process seems not to need assistance from her intention in order to get where it's going, chugging along as it does according to principles that appear to have very little in common with ordinary psychological ones. The presumed self-sufficiency of physiological processes can, in a similar fashion, appear to muscle psychological states quite generally out of the causal picture.

Some philosophers hold that when we take a steely-eyed look at this state of affairs, we are forced to admit that common sense is simply mistaken about why people behave as they do. Others seek to maintain our common-sense understanding of ourselves, and attempt to show that the two kinds of causal explanation are in fact compatible. But how can this be achieved? Here is one proposal: We ought to allow that there is a systematic overdetermination of events by their causes. To revert to the example above, we ought to allow that the rising of the woman's hand is caused by both her physiological state *and* her psychological state. In the philosophy of mind, however, there

is significant resistance to accepting any sort of systematic overdetermination. This is in part because the concept of overdetermination is commonly thought—erroneously, I will argue—to require that each state be metaphysically independent of the other; and systematic overdetermination of *that* sort would surely be wildly implausible. But resistance to allowing overdetermination runs deeper. In this paper, I look into two of the deeper forms refusal to countenance overdetermination has taken, and show how each threatens to undermine the view that there are any mental causes at all. The first, underlying 'the steely-eyed' look, explicitly rules out the possibility of reconciling the manifest image and the scientific image of the world from the start.[1] This principle, the Strong Principle of Non-Overdetermination, will be the topic of section one. The second underlies the widely-held view that the mind must be 'naturalized' to be rendered compatible with the natural sciences, and leads to the worry that the mind cannot be understood naturalistically. I call this principle the Weak Principle of Non-Overdetermination, and discuss it in section two. In section three, I will call both of these principles into question and show that we can conceive of a sort of overdetermination that, far from being wildly implausible, accords well with our common-sense view of the world.[2]

More broadly, the aim of this paper is to negotiate a dialectical impasse in the philosophy of mind. According to naturalists, we should not include in our ontology anything but what the natural sciences equip us to describe. Some naturalists are optimistic about the prospects that the mind will pass this test, others are not. I'll call the former *naturalist realists* and the latter *eliminativists*. Much of naturalism's appeal consists in the fact that it looks like the only alternative to what I will call *unnaturalism*. Unnaturalist views are those that postulate some sort of immaterial soul to account for human behavior. I will take it for granted that such views are mistaken. But if unnaturalism were the only alternative to naturalism, then we would be stuck either pursuing the heretofore unsuccessful project of naturalizing the mind or having to accept the incredible claim that there really are no minds after all. This reckoning of matters, however, overlooks the possibility of a third way: the view that the mind is both real *and* unnaturalizable. Such a view would be

---

1 The phrases "manifest image" and "scientific image" are, famously, taken from "Philosophy and the Scientific Image of Man" in Wilfrid Sellars, *Science, Perception, and Reality* (London: Routledge & Kegan Paul Ltd., 1963).

2 Two recent attempts to solve the problem of mental causation by accepting some form of systematic overdetermination (albeit systematic overdetermination conceived differently than I do here) are Eugene Mills, "Interaction and Overdetermination", *American Philosophical Quarterly*, vol. 33, no. 1 (1996) and Brian Jonathan Garrett, "Pluralism, Causation, and Overdetermination", *Synthese*, vol. 116 (1998). Whereas Mills argues in favor of event-dualism, and Garrett defends a form of event-pluralism, I here remain neutral in what follows on the question of whether mental events are token-identical to physical events.

non-naturalist without being *unnaturalist*. I will call it *naive realism*. The burden of this paper is to argue that this view is coherent, and that a proper appreciation of its coherence shifts the burden of proof in the contemporary debate over mental causation in the philosophy of mind.

Naive realism, however, will not appear to be an option for those who hold either of the two principles of non-overdetermination. More narrowly, the point of this paper is thus to show that there is a sensible way to relax the metaphysical interdiction on allowing overdetermination, one that does not require that we accept anything supernatural or mystical. In section three, I show how this can be done, and indicate how we can think of mental causation unnaturalized but *not* unnatural. My argument there is not designed to refute naturalism as an empirical hypothesis about what success the natural sciences can have in explaining a domain that has so far proved elusive. Rather, it is designed to show that there is no metaphysical obstacle to abandoning that empirical hypothesis while still rightfully retaining our belief in the mind's efficacy and reality.

§1 The Strong Principle of Non-Overdetermination (SP)

In this section, I review one sort of argument that has often been made in support of epiphenomenalism, the view that the mind is causally inert, and show how it depends on finding a certain kind of overdetermination unacceptable. In §1.1, I say what the Strong Principle of Non-Overdetermination is, and indicate that it can be formulated with some clarity. In §1.2, I show how SP can be used to establish that almost nothing we believe about the world is true. In §1.3, I underscore how serious these consequences are.

§1.1 *The Content of The Strong Principle of Non-Overdetermination.* The point of this section is to offer a preliminary formulation of a principle that is presupposed by a variety of arguments that have been advanced in favor of epiphenomenalism, and to show how it can lead to a troublesome and articulate skepticism about mental causation.[3] Ultimately, I will be concerned to

---

[3] This particular source of epiphenomenalist anxiety has been discussed as 'the problem of exclusion'. See Jaegwon Kim, *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993), especially essay 13, Ernest Lepore and Barry Loewer, "Mind Matters", *Journal of Philosophy*, vol. 84 (1987), "More on Making the Mind Matter", *Philosophical Topics* vol. XVII (1989), Jerry Fodor, "Making the Mind Matter More", *Philosophical Topics* vol. XVII (1989), Mark Johnston, "Why Having a Mind Matters" in Ernest Lepore and Brian McLaughlin, *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (New York, NY: B. Blackwell, 1985), Ned Block, "Can the Mind Change the World?" in Cynthia Macdonald, *Philosophy of Psychology: Debates on Psychological Explanation* (Cambridge: Basil Blackwell, 1995), Stephen Yablo, "Mental Causation", *The Philosophical Review*, vol. 101, no. 2 (1992), Robert Van Gulick, "Three Bad Arguments for Intentional Property Epiphenomenalism", *Erkenntnis*, vol. 36, no. 3 (1992), and David Robb, "The Properties of Mental Causation" in *The Philosophical*

argue against this principle; but it is crucial to see first that it can be formulated consistently.

> *The Strong Principle of Non-Overdetermination* (SP): If an event e has a complete causal history at a causal level l, then there are no non-l causal histories of e and no non-l property can be efficacious in relation to e.

We can begin the process of saying what this means by introducing the idea of a causal level. This notion is implicit in most of the writings on this topic and in ordinary conversation, where we know the difference between looking for the cause of some event, a homicide, for example, at the level of physiology (what was the 'cause of death'?), psychology (who had a motive?), sociology (what's wrong with our country?) or evolutionary biology (how did our species become capable of such heinous acts?). Accordingly, we can identify a causal level with the domain of a discipline that explains why events unfold in the world the way they do. So we can speak of the level of physiology, psychology, sociology, etc.[4] Even if every spatiotemporally existing thing is composed of fundamental physical entities, these entities are often organized into such groups as are describable in terms that are not those of physical laws. When this is so, we can speak of there being a level of objects, states, and events corresponding to their non-physical descriptions, and of specific causal principles that govern their relations. (So every level l corresponds to a canonical l-language.) Adopting some such view is the only way of making sense of a universe that (ostensibly) contains such diverse kinds of things as planets, countries, corporations, people, livers, viruses, molecules, and electrons.

Next we should look at the notion of a causal history:

Quarterly, vol. 47, no. 187 (1997). It has been the source of the widespread objections raised to Davidson's argument for anomalous monism in "Mental Events", from *Essays on Actions and Events* (New York: Oxford University Press, 1980). For examples of these criticisms, see Fred Stoutland, "Davidson on Intentional Behavior" in Lepore and McLaughlin 1985, Ted Honderich, "The Argument for Anomalous Monism", *Analysis*, vol. 42, no. 1 (1982) and Ernest Sosa, "Mind-Body Interaction and Supervenient Causation", *Midwest Studies in Philosophy*, 9 (1984).

4 There have been some rigorous attempts to make the notion of a level precise, to answer questions about where one level starts and another begins, but these accounts generally incorporate assumptions about what relations between such levels must be like that it is part of the aim of this paper to dispute. See, for example, William Wimsatt, "Forms of Aggregativity" in ed. Alan Donagan, A. N. Perovich Jr., and M. V. Wedin, *Human Nature and Natural Knowledge* (Boston: D. Reidel Pub. Co., 1986) and P. S. Churchland and Terrence Sejnowski, "Brain and Cognition" in Michael Posner, *Foundations of Cognitive Science* (Cambridge, Mass.: MIT Press, 1989).

> The *causal history* at level l of an event e at l' is the set of events at l that stand in causal relations with e, by the standards appropriate to l and l'.

To the extent that the causal relation is, quite generally, a transitive relation, a causal history of an event e at l will include all of the causes (at l) of e's causes. Thus, the physical causal history of a physical event might, on this conception, extend as far back as the beginning of the universe.

It might be thought that my attempt to formulate SP should necessarily take a detour into a general theory of causation that would fit with my use of causal terms in this definition. But no such detour is necessary. The only presupposition of the conception of causal histories employed here is that there are *some* standards proper to each causal level that cases of causation must meet in order to count as causes at that level. And it is part of what it is to have an understanding of a domain at all that one have at least a rough idea of why things happen the way they do in it. So I mean to leave it open how the standards and characteristics of intra- and inter-domain causal relations between events will vary from domain to domain. SP is to be understood as a proposed *constraint* on any conception of causation; it is put forward as a basic part of our understanding of how the world works.

Finally,

> A *complete causal history* at l of an event e is a causal history h of e where causal relations between events in h and between the events in h and e can be fully accounted for using canonical l-vocabulary.

If there is a level l where every event has a complete causal history, then the complete causal history h (at l) of an event e will include complete causal histories (at l) of all of the events in h.

So far we've talked only of causation between events, but SP also places a restriction on properties. This clause is necessary in order to show that the epiphenomenalist concern arises whether or not one accepts the thesis that mental events are token-identical to physical events.[5] A property is causal if and only if it is one in virtue of which an event can have effects of some kind or another, where 'in virtue of' introduces any aspect of an event that is supposed to answer the following kind of question: "What was it about event c that causally explains its having effect e?"[6]

Now that we know what SP says, I will go on to showing how devastating to our conception of the world it would be were we to accept it. The

---

5 Davidson dismisses this concern in an uncharacteristically careless way. See Davidson, "Thinking Causes" in *Mental Causation*, ed. John Heil and Alfred Mele (New York: Oxford University Press, 1991).

6 See Kim, "Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism?" in Heil and Mele 1991 for a discussion of this point.

intuition behind SP is that the causal *completeness* of physical causal systems (the fact that one does not, in principle, need to refer to other kinds of properties to account for what happens in them) entails the causal *closure* of physical causal systems (the idea that physical events do not causally interact with non-physical events or with physical events in virtue of their non-physical properties).[7] This intuition has often been exploited to cast doubt on the idea that mental events in particular are not efficacious *qua* mental. But this intuition has more general significance. We will see in the next section how it can be used to cast doubt on the idea that there are any real causal properties at all beyond physical properties.

§1.2 *Consequences of SP.* In this section, I show that SP has eliminativist consequences few philosophers would want to accept. I will consider two variations on a basic argument scheme, corresponding to the two positions on whether particular high-level events (of which mental events are an instance) can be identified with particular physical events. This pair of arguments shows that we can derive precisely the same conclusions from either the assertion or the denial of token-identities between mental and physical events. In my view, the important issue is not whether we should accept the doctrine of token-identity, but rather whether we can allow the sort of overdetermination that SP rules out. If we accept SP, both positions are susceptible to arguments requiring that we reject mental causation; if we reject SP, both positions will leave realism about mental causation a viable option.

The first argument showing that SP requires eliminativism about every level other than physics goes as follows: If one accepts the view that high-level events are *not* token identical to any physical events, then, one ought to accept the following argument. From these premises,

(1) SP: If an event e has a complete causal history at a causal level l, then there are no non-l causal histories of e and no non-l property can be efficacious in relation to e.

(2a) *The Generalized Principle of Causal Interaction for Events*: All events have physical effects; and[8]

---

7 The appeal of this inference has lead some to reject the idea that physical causal systems are causally complete. See for example Lynne Rudder Baker, "Metaphysics and Mental Causation" in Heil and Mele 1991, and E.J. Lowe, "The Problem of Psychophysical Causation", *Australian Journal of Philosophy*, vol. 70, no. 3 (1992). If my argument here is successful however, this inference is a bad one and the causal completeness of physical systems, even if correct, is harmless.

8 This is a variation on Davidson's "Principle of Causal Interaction" from his argument for token-identity. See his "Mental Events" in Davidson 1980. I will not undertake a defense of (2a) here, but I think it is difficult to deny. A hurricane can't blow a house down if it

(3) *The Principle of Causal Completeness of the Physical* (CCP): All physical events[9] have complete physical causal histories.[10]

we can infer the non-existence of non-physical events. By (3), any physical event has a complete physical causal history. By (2a), every event has physical effects. But by (1), a physical event has no causal histories other than the physical one. So there are no non-physical events. To deny that there are non-physical events, however, is to deny that non-physical properties are ever instantiated. We can conclude that, if our conception of differing causal levels commits us to rejecting the token-identity, it thereby also commits us to the rejection of an ontology that includes levels higher than the physical.

The second argument showing that SP requires eliminativism about every level other than the physical has the same form. If one accepts the view that high-level events *are* token identical to physical events, then, one ought to accept the following argument. From these premises,

(1) SP;

---

can't move molecules; if my deciding to pull the trigger couldn't affect the atoms composing the trigger, how could I ever shoot someone?

9 Throughout this paper, I use "physical" and its cognates as variations on "physics". There is still a thorny question, however, about what exactly a physical property is and how, precisely, physical properties are to be distinguished from mental properties. In "There is No Question of Physicalism", *Mind* vol. 99 (1990), Tim Crane and D.H. Mellor argue that "there is no divide between the mental and the non-mental to set physicalism up as a serious question" (206), and thus also no way to make a principle such as CCP meaningful. (More recently, Barbara Montero has argued for a similar view in "The Body Problem", *Noûs* vol. 33, no. 2 (1999).) If this claim is correct, then the problem of mental causation simply does not get off the ground. I will assume, however (as Crane himself does in his "Mental Causation", *Proceedings of the Aristotelian Society*, supp. (1995)), that there is some way of saying what is meant by 'physics', and that mental properties will not fall into the domain of physics, so understood. My argument here can thus be understood as having the following form: *Even if* CCP is both meaningful and true, there is no problem of mental causation.

10 This does *not* mean that every physically describable state of affairs is causally determined by prior physical states of affairs, nor does it mean that every feature of every physical event has a cause. Famously, there are significant correlations between the goings-on in spatially disparate quantum mechanical systems that have no common cause. It only means that one need never depart from the physical level to account for what happens there. The just-mentioned correlations would only violate the causal completeness of the physical if there were some reason to think that they were the result of higher-level intervention. It may be that there are some physical events that have no causes at all. If so, we can modify CCP slightly while leaving these arguments intact. The modified version might go: All physical events either have complete physical causal histories or have no causal histories at all.

(2b) *The Generalized Principle of Causal Interaction for Properties*: Every property is one in virtue of which an event can have (physical) effects; and[11]

(3) CCP;

we can derive the non-existence of high-level properties. By (3) any physical event has a complete physical causal history. By (2b), any non-physical causal property would be one in virtue of which an event can have physical effects. But by (1), no non-physical property can be efficacious in relation to a physical event. So no non-physical properties are ever actually instantiated. We can conclude that, if our conception of differing causal levels commits us to accepting token-identity, then there are no real non-physical properties, hence no true non-physical characterizations of events.

If SP is true then, regardless of whether we think of mental events as token-identical to physical events, we are required to be eliminativists about almost everything. There are no psychological states (as there could be no events of entering into such states), or for that matter any people. There are no buildings or bridges, no chairs or tables, no avalanches or hurricanes, no insects or animals, and so on. The only things that exist are what true physical theories say exist.

§1.3 *The Seriousness of SP's Consequences.* There are a variety of strategies a proponent of SP might deploy at this point to blunt its impact. One would be to object to the final inference in the second argument of the previous section—from the claim that there are no real non-physical properties to the conclusion that there are no true non-physical characterizations of events. It has been argued by some that a predicate might apply to an object without there being a property both expressed by the predicate and possessed by the object.[12] An advocate of SP might attempt to exploit this point in the

---

[11] (2b) is plausible for the same reasons as (2a). Just as real events must be able to make a difference to what happens, real properties must also be those that can make a difference to what happens. (And in this variation of the argument, monism is true by hypothesis, so making a difference to what happens means making a difference to physical states of affairs.) This constraint adds that the sort of difference real properties must be able to make is a difference made by their properties *as such*, and not in virtue of their association with or supervenience on properties that do make a difference as such. To use a popular example: A shot killed Dillinger. The shot was loud, but it was not its loudness that killed him. (2b) says: If loudness were always only as efficacious in relation to the effects of loud events as it is to this shooting death, then we would have no reason to believe that loudness was a property that really belongs to events in the world. I should also note that this principle is meant to cover only the sorts of properties that can be acquired or lost—the inefficacy of atemporal properties, such as being an odd number, should be thought to detract neither from their reality nor from the plausibility of (2b).

[12] See, for example, D. M. Armstrong, *A Theory of Universals* (Cambridge: Cambridge University Press, 1980), Sydney Shoemaker, "Causality and Properties" in *Time and*

following way. Although there are no mental properties, and also no mental events distinct from physical events, there may still be physical events (i.e., events with physical properties) that can truly be described in mental terms. Hence, no eliminativism.[13]

It should be noted, first, that this move would be of little help in solving the problem of mental causation. The epiphenomenalist who accepts token-identity holds that although some physical events have mental properties, it is not in virtue of a physical event's mental properties that it has the effects it does. (He thus rejects 2b.) But such a position is not a way of solving the problem of mental causation; rather it is a way of accepting that the mind is not efficacious as such. And a philosopher who holds that physical events do not even *have* mental properties goes one step further than the epiphenomenalist. In this case, it is difficult even to know what to make of the question of whether an event has effects in virtue of *being* mental—it certainly does not have effects in virtue of its mental properties, for it has none.

Nonetheless, in arguing that SP entails eliminativism I do assume that the following principle is true:

(P): Where 'F' is a predicate and 'a' is an object or event, 'Fa' is true only if the object or event designated by 'a' has the property expressed by 'F'.

Though a satisfactory defense of P would take us well beyond the scope of this paper, I will respond briefly to what appears to me to be the most plausible rationale for rejecting it. Citing Wittgenstein, D.M. Armstrong notes that the predicate 'is a game' does not apply to games in virtue of a single property all games share; John Heil makes a similar case for 'is a stone'.[14] This is no reason to deny that there are any games (or stones), these authors contend; it is rather a reason to reject the idea that this predicate applies in virtue of a property it expresses and that is possessed by all and only those objects in the predicate's extension.

But the Wittgensteinian consideration in fact does not support this conclusion. In my view, the right lesson to draw from it is rather that the following two claims are consistent: (1) Each game has the property of being a game. (2) There is no *further* property that all games have in common. There is no one property possessed by every game in virtue of which it counts as a game—but this does not require that we give up on the sensible idea that 'is a

---

Cause, ed. Peter Van Inwagen (Boston: D. Reidel Pub. Co., 1980), and John Heil, *Philosophy of Mind : A Contemporary Introduction* (New York: Routledge, 1998).

13 This problem was pointed out to me by an anonymous referee at this journal.

14 See Armstrong, "Towards a Theory of Properties: Work in Progress on the Problem of Universals", *Philosophy* 50 (1975), pp. 149–50, Heil 1998, 194–97 and Heil, "Multiple Realizability", *American Philosophical Quarterly*, vol. 36, no. 3 (1999).

game' designates a property possessed by (and only by) every game. The fact that different games may bear only a family resemblance to one another is thus not a blow to the ontological status of games; the idea that nothing possesses the property of being a game, I contend, would be.[15]

It is also worth pointing out that we need only resort to the denial of the above principle to save realism about mental discourse if we have grounds for privileging physical predicates over all others. Typically such grounds are sought in the nature of causation. But it is the contention of this paper that the nature of causation provides no such grounds.

Consider next the philosopher who accepts SP and consoles us with the idea that high-level causal claims, while not true strictly speaking, are true in some weaker sense—in the sense of being psychologically inescapable for us, or in some other way practically indispensable. The thought might, for example, continue this way. "In our everyday dealings with the world, we can't help but conceive of the world as containing more than it strictly speaking does—we can't help but conceive of it *as if* it contained all manner of things beyond the physical. It is the practical indispensability of our belief in this as-if world that legitimizes our continuing to talk and think in the ways we always have. Nevertheless, in the seminar room, when we abandon the practical standpoint and seek to understand the world as it is in itself, we should admit that the high-level entities of this as-if world do not really exist; that is, we should admit that the as-if world is just a projection of our common-sense conceptual schemes."

I don't find this line of thought, which might be called 'seminar naturalism', particularly consoling. I can imagine someone who is alarmed by SP's most radical consequences still feeling prepared to embrace seminar naturalism with regard to the cases of airplanes, heart medicine, rifles, and any other where our concerns are largely instrumental—where we are mainly interested in, say, arriving at our destination, staying alive, and snagging a few deer. If it is not true, strictly speaking, that my arrow killed the doe, it doesn't make the venison any less tasty. But for claims to be less than true, strictly speaking, is for claims to be false; and the idea that there really aren't any actions or any psychological events is genuinely disturbing. And this is why we have anthologies devoted to concerns over mental causation, and not meteorological causation, despite the fact that the intuitions that fuel those concerns sometimes apply equally well to all non-fundamental domains. But if we could allow that there *are* violations of SP, then we would not need the

---

[15] In "The Status of Content", *The Philosophical Review* vol. 99, no. 2 (1990), Paul Boghossian comes to a similar conclusion: "If there are extra-linguistic psychological properties for the sentences of physics to answer to, but no extra-linguistic properties for the sentences of psychology to answer to then it isn't true, in the strict and literal sense, that there are *true* sentences of psychology" (pp. 179–80).

consolations of the seminar naturalist. And I will argue that we can, therefore we don't.

§2 The Weak Principle of Non-Overdetermination (WP)

In this section, I want to understand SP in relation to another principle, the weak principle of Non-Overdetermination. WP is an alternative response to the suspicions about overdetermination that are expressed by SP. According to WP, there are always mechanisms that mediate higher-level causal relations. I will begin in §2.1 by showing how the concept of a mechanism is relied on by two mainstream naturalist philosophers of mind: Jerry Fodor and Terence Horgan. I will show that commitment to seeing all high-level causation as mediated by mechanisms is more substantial than it is sometimes taken to be. I will then, in §2.2, describe what I take to be the dialectical connections between SP and WP. This will set up the next section, which is devoted to casting doubt on both principles and the underlying intuition that fuels them.

§2.1 *Invocations of 'Implementing Mechanisms'*. The concept of an implementing mechanism is invoked by many mainstream naturalists. Horgan (a self-proclaimed "metaphysical naturalist") offers four "inter-level constraints" as part of articulating his "physicalist metaphysical *Weltanschauung*." [16] The third states:

> For any causal transaction where some higher-level property F is cited as causally explaining the effect, there must be an *underlying mechanism* in virtue of which the transaction occurs—a mechanism involving a physical property (or complex of physical properties) which, on the given occasion, *physically realizes* the property F. That is to say, causal transactions invoking higher-order properties must be grounded in causal mechanisms involving the nexus of physical causes and effects, mechanisms describable and explainable at the level of physics.[17]

Fodor accepts the very same explanatory burden, saying that we should accept "functionally defined theoretical constructs only where mechanisms exist that can carry out the function and only where [we have] some notion of what such mechanisms might be like."[18] This burden is a consequence of his view that "P is a causally responsible property if it's a property in virtue of which individuals are subsumed by causal laws,"[19] and that "non-basic [causal] laws require mediation by intervening mechanisms."[20] The distinction

---

16 Terence Horgan, "Non-Reductive Materialism and Explanatory Autonomy of Psychology" p. 301 in ed. Steven Wagner and Richard Warner, *Naturalism: A Critical Appraisal* (Notre Dame, Indiana: University of Notre Dame, 1993).

17 Ibid. 302, his emphasis.

18 Jerry Fodor, "The Mind-Body Problem", p. 34 in ed. Richard Warner and Tadeusz Szubka, *The Mind-Body Problem* (Cambridge, Mass: Blackwell, 1994).

19 Fodor 1989.

20 Ibid., p. 74.

between basic and non-basic laws is just the distinction between fundamental and non-fundamental laws, and Fodor cashes it out as follows:

> [A] metaphysically interesting difference between basic and nonbasic laws is that, in the case of the latter but not the former, there always has to be *a mechanism in virtue of which* the satisfaction of its antecedent brings about the satisfaction of its consequent. If 'Fs cause Gs' is basic, then there is no answer to the question *how* do Fs cause Gs; they just do....[21]

Fodor does not simply take it for granted that "all mechanisms that mediate the operations of laws are eventually physical". Rather he offers it, facetiously, as "a bold assumption" (i.e., a trivial truth), one he distances himself from only "because [he doesn't] know what it *is* for a mechanism to be physical as opposed to spiritual." In a footnote he expands on the idea:

> 'Eventually' means: either the law is implemented by a physical mechanism, or its implementation depends on a lower-level law which is itself either implemented by a physical mechanism or is dependent on a still lower law which itself is either implemented by a physical mechanism or...etc. Since only finite chains of implementation are allowed you have to get to a physical mechanism 'eventually'.... And though, presumably, physical mechanisms implement every high-level law, they usually do so via lots of levels of intermediate laws and implementation.[22]

I think a close reading of many naturalist philosophers would uncover a very similar commitment.

But I don't think the notion of an *implementing mechanism* is a harmless one, Fodor's nonchalance notwithstanding. The question of what a mechanism is exactly, and the further question of what distinguishes mechanistic explanations from other kinds are deep and interesting ones for the philosophy of science. Clearly this is not the place to answer them, but I do want to formulate, in a very general way, what I take to be the upshot of this mechanistic constraint.

According to The Weak Principle of Non-Overdetermination, the causal powers of any high-level object, event, property, or fact can always be completely accounted for by reference to the causal powers of its internal lower-level components and the causal significance of their mechanistically describable connections. The adoption of this constraint results in the view that the world can be understood as a mechanistic hierarchy of causal levels (what I will call *a mechanistic world-view*). I will call explanations that perform such reductions *lower-level mechanistic explanations*.[23]

---

21 Ibid., p. 66, his emphasis.

22 Ibid., p. 76 and footnote.

23 See William Bechtel and Robert Richardson, *Discovering Complexity* (Princeton: Princeton University Press, 1993), for one way of working-out the details of what I call 'lower-level mechanistic explanations'.

> *The Weak Principle of Non-Overdetermination* (WP): Every legitimate non-physical causal claim presupposes the existence of some lower-level mechanistic explanation.[24]

According to the world-view that motivates this principle, part of what makes physics the *fundamental* level is that it is only true for what happens at this level that no lower-level mechanistic explanations are possible.

Some naturalists create the impression that WP is no more controversial than the view that there are no supernatural forces at work in the world. In studying the mind, naturalists offer WP as a way of putting the perfectly sensible thought that there must be some accounting for what we are capable of in terms of the way our bodies are constructed. It is part of the aim of this paper to show that there is a significant gap between affirming this perfectly sensible thought and WP.[25]

For the present, it suffices to note that WP is substantial enough to threaten our most deeply held views about ourselves and the world. For example, to many, the idea that connectionist systems are accurate models of the etiology of human behavior has eliminativist ramifications. According to WP, if a connectionist story (a description of the units that make up a network, the specifics of the 'subsymbolic' connections between them, and the 'learning' algorithm governing changes in the weights of those connections) accurately represents those aspects of our brains' machinery specifically relevant to our capacity to think, and mental states can't be plausibly identified with elements in these models, then there is no room left for them in our understanding of the causal order.[26] The difference between an eliminativist (of one sort, anyway) and a naturalist realist is that the former believes mental causation will not meet WP, and the latter thinks it will. According to naive realism, mental causation need not satisfy WP in order to prove its legitimacy.

---

[24] Of course there are many variations in the class of position I'm trying to single out. Lepore and Loewer reveal their allegiance to WP when they treat it as a blow to the causal efficacy of "content" properties that there are no "physicalistic explanations" of their causal powers. Their project is to find some weaker notion of causal relevance with which we might console ourselves in the absence of full-bodied causal potency. See Lepore and Loewer 1989. In a different vein, Dretske sees himself as rescuing mental causation when he argues that there *are*, in a sense, physicalist explanations of the causal powers of mental states. See Dretske, "Mental Events as Structuring Causes of Behavior" in Heil and Mele 1991. Other approaches include one pursued independently by Cynthia and Graham Macdonald, "Mental Causation and Explanation of Action" in ed. Leslie Stevenson, Roger Squires and John Haldane, *Mind, Causation and Action* (Oxford: Basil Blackwell, 1986), and in Stephen Yablo 1992.

[25] I will explicitly return to this gap again in §3.4.

[26] See, for example, Paul Churchland, "Eliminative Materialism" in *Matter and Consciousness* (Cambridge, Mass.: MIT Press, 1984), and William Ramsey, Stephen Stich, and Joseph Garon, "Connectionism, Eliminativism, and the Future of Folk Psychology" in ed. John Haugeland, *Mind Design II* (Cambridge, Mass: MIT Press, 1997).

Now that we have a sense for what WP says, and why it is important, I want to consider it in relation to SP.

§2.2 *Dialectical Relations between SP and WP.* In this subsection I will look at the dialectical relationship between SP and WP, which is somewhat complex. On the one hand, thoughtful consideration of the rationale behind WP might lead one to give it up, and adopt SP instead. On the other hand, WP can look like a safe fall-back position for someone who sympathizes with the fear of overdetermination that seeks to express itself in SP, but who finds SP's consequences unpalatable. This dialectical stage-setting prepares us for the argument of the next section, in which both principles are rejected.

Let us first consider WP as a step in the direction of SP. The view that one can always give a lower-level mechanistic explanation for any particular case of non-fundamental causation amounts to the idea that non-fundamental causation can equally well be viewed at a lower level. For in giving a lower-level mechanistic explanation, one is exhibiting that in which, in a given kind of case, or on a particular occasion, high-level causation consists. Causation is aggregative in the sense that emerged in the previous section: The causal powers of any high-level entity can be divided up and apportioned out to the various lower-level components of the system and their lower-level connections without remainder. We can speak of the causal power of the higher-level entity, but the validity of such talk rests on the in-principle possibility of such an apportioning.

'Without remainder' thus signals a threat implicit in a certain way of cashing out the distinction between spurious and real causal properties. It has been argued that lower-level causal stories vindicate high-level causal claims because they show *how* the high-level cause brought about its effect.[27] On this conception, in giving a lower-level explanation of the very same event, we learn from where a high-level process derived its causal power. And if no lower-level account is in the offing, then we know we are dealing with a spurious causal process, with a pseudo-process.

'Without remainder' also suggests a connection between WP and SP. According to WP, all high-level causation reduces to (consists in) fundamental causation, so every causal history of an event must in that sense be part of the one complete and fundamental causal history of that event. It is the ease with which one can slide from the thought that *all causation is ultimately physical* (WP) to the claim that *there is only physical causation* (SP) that partially accounts for how someone might come to believe SP. If one thinks that physics can vindicate high level causal claims, but not the other way around, then it would not be hard to come to view the physical as where it's *really* at, causally speaking. One might thereby arrive at the view that only

[27] See David Henderson, "Accounting for Macro-Level Causation", *Synthese* 101 (1994).

the physical is the realm of the genuinely causal. Mechanistically understandable high-level claims might still appear more tolerable than other sorts of high-level claims, as they can at least be backed up in a special way by the causal account at the fundamental physical level. Nonetheless, they would on this view still be false (or only 'approximately' or 'metaphorically' true). One might thus reason from WP to SP.

But there is also a way that someone might arrive at WP by way of SP. (This is not to say that WP is motivated solely by SP. We will look at other factors in §3.) Faced with the unpleasant ramifications of SP, but unwilling to surrender the underlying intuition, someone might withdraw to WP. WP can be represented as a simple modification of SP. To

> If an event e has a complete causal history h at a causal level l, then there are no non-l causal histories of e and no non-l property can be efficacious in relation to e,

we can add

> except those causal histories and properties that can be mechanistically explained in terms of h.

This new clause need not be too bitter a pill to swallow for someone anxious about overdetermination, for mechanistically explainable causation *is* ultimately physical causation. As such, it represents only a modest departure from SP and appeals to the same intuition. An advocate of WP still refuses to acknowledge the possibility of causation that doesn't consist in physical causation—so no scruples about overdetermination will be offended. This is why WP can look like a safe fall-back principle for someone initially inclined towards SP, but squeamish about the wild eliminativism canvassed in §1.2. Such a philosopher might think, having accepted WP as an expression of the same intuition that underlies SP, that it captures all that's right about SP. WP is a more forgiving expression of the intuition behind SP; it alone does not have the consequence that all high-level causal claims are false. It just limits the high-level claims that can be true to those that can be given a lower-level mechanistic explanation; it limits causal properties to those that figure in mechanistically-understandable special sciences.

We are thus now in a position to notice a telling instability. Both WP and SP have as a consequence that every instance of causation can be represented at the fundamental level. SP concludes that all high-level claims are false. WP insists, the failure of type-reductions between sciences notwithstanding, that high-level causal claims are true iff they can be understood as a different, albeit less perspicuous, means of exhibiting that fundamental causal level. But a proponent of SP does not deny that we can make a significant distinction between high-level causal claims that are mechanistically understandable

and those that are not; he just insists that they are all, strictly speaking, false. Both views endorse a metaphysical hierarchy with the following structure:

Fundamental Physical Properties
Mechanistic Properties (e.g., aeronautical, geological)
Useful Non-Mechanistic Properties (e.g., mental, economic)[28]
Useless Non-Mechanistic Properties (e.g., astrological, numerological)

An advocate of SP would say that the metaphysically significant distinction that needs to be drawn here is directly under 'Fundamental Physical Properties'; an advocate of WP wants to locate the metaphysically significant distinction directly under 'Mechanistic Properties'. The rationale for the fall-back position, as opposed to a view that does not privilege mechanisms in this way at all, is the vindicating function assigned to fundamental physical causation. The commitment to placing physical properties higher on this table than mechanistic properties is a necessary consequence of this position. But the vindicating power of the physical is difficult to maintain without according it ontological privilege. The rationale behind WP thus still suggests that the real causal work is being done at the physical level. Hence the instability: One flees from the incredible consequences of SP in hopes of satisfying the underlying intuition in a less destructive way. But in WP, one ends up with a half-hearted expression of the intuition whose resolute expression entails the damaging ramifications detailed in §1.2.

A central concern of this paper is to recommend that we reject the idea that *all* causation must be ultimately understandable as fundamental causation, that we reject the view that either the first or the second line separates the real from the illusory. According to naive realism, the only metaphysically important dividing line is the third. This is what distinguishes the naive realists from the naturalists, and it is also what makes a robust realism about the manifest image of the world a viable option. The claim that there can be causation that is *not* ultimately physical is a general metaphysical thesis with special importance for the philosophy of mind. It has special importance here because if we accept it, we will be even less tempted by the eliminativism of SP and at the same time freed from epiphenomenalist concerns generated by the fear that the high-level descriptions to which we are most attached will not meet WP.

---

[28] This is a tendentious category. In §3, I will provide a rationale for thinking that mental properties belong on this list. A different argument would be required to show that economic properties do as well. I have this row here to indicate how a proponent of WP or SP would be required to view such properties, if she could be persuaded that there were any.

Both SP and WP are motivated by the same intuition. I don't think this is at odds with the fact that they are also strongly independent. Indeed, one needs to choose between them as the best expression of a worry about overdetermination, for they are in tension with one another, differing as they do on the status of mechanistically understandable high-level causation. Thus I do not think that naturalists hold both SP and WP. My claim is not that most naturalists first take the trouble clearly to distinguish SP and WP and then somehow manage to commit themselves to both. My claim is rather that both principles stem from a common concern, and that they are dialectically related in the following way: Each principle is unsatisfying in a respect in which the other is not. WP appears to preserve a place for high-level causation, but insofar as it is still committed to ultimately identifying all such causation with physical causation, this allowance inevitably rings somewhat hollow. SP is more faithful to the underlying intuition, but only at the cost of outrageous consequences. I conjecture that one source of the appeal of a certain range of positions in the philosophy of mind is that the pervasively unarticulated ground of their opposition to overdetermination waffles between commitment to SP and commitment to WP, thus engendering the illusion of the possibility of occupying a position in the philosophy of mind which simultaneously retains the attractive aspects of SP and WP, while simultaneously eschewing the unattractive consequences of each.

§3 Beyond SP and WP

In this section I will sketch a rationale for overcoming the suspicion of overdetermination expressed in SP and WP. I will offer reasons for thinking WP is too strong (i.e., excludes too much), from which it follows that SP (which excludes even more) is also too strong. §3.1 shows that the sort of overdetermination that naive realism requires is not metaphysically spooky. In §3.2, I give some reasons for thinking that our understanding of mental causation has very little to do with mechanistic causation, and should thus not be measured by its capacity to satisfy WP. And in §3.3, I will argue that this view does not make a mystery of the place of the mind in the physical world.

§3.1 *Genuinely Spooky Overdetermination.* Insofar as this paper is an attempt to support naive realism, I am advocating that we allow a certain kind of systematic overdetermination of events by their causes. Say, on a whim, I drive my golf-cart through a sliding glass door. There is some physical event corresponding to the breaking of the window. Like all physical events, it has a complete causal history at the level of physics. But it also has causes at other levels as well. At the level of mechanical engineering, there is a causal history involving the propulsion of the cart towards the window. At the level of the mechanics of middle-sized objects, there is a

causal history involving a heavy, swiftly moving body striking a fragile surface. At the level of psychology, there is causal history involving my destructive and capricious personality, etc. What distinguishes naive from naturalist realism is the belief that some of the causal histories might be genuine yet *not* be mechanistically understandable.

One explanation for the squeamishness some philosophers feel at this is that there is a *kind* of systematic overdetermination the idea of which *should* strike us as quite spooky. If my alarm clock failed to go off, this might be because I forgot to set it, or because the power went out during the night, or because someone else turned it off after I had set it, and so on. In ordinary circumstances, we would be surprised to find, for example, that the power went out *and* that the clock was defective. We would think it a suspicious coincidence or just a fluke. Both in daily life and in scientific practice, there is often a presumption that two *genuinely* different causal accounts of the same event cannot both be true. If every high-level cause of a physical event e entails the existence of multiple *independent* causal chains, each of which is *sufficient* for e, then we would have good reason to be spooked. Sometimes events are overdetermined in this sense, of course. But a case of this sort is always either the result of coincidence or the product of design. Occasionally it just happens that, for example, two hunters independently shoot the same deer at the same time, where either shot would have killed it. We can call such cases overdetermination by coincidence. In other cases, some person or organization might ensure that some event occurs by setting up several causal chains which result in that same event. We can call this sort of case overdetermination by design. The thought that overdetermination by independent causal chains is routine, however, would offend the sensibilities of philosophers and non-philosophers alike. We would demand an explanation of the ubiquitous co-occurrence of such causal chains. And nothing short of finding a common cause or hidden dependence of some of these causal chains on others would satisfy this explanatory demand.

But the overdetermination required by the naive realist is not of this sort. High-level causal processes are *always* dependent on physical causal processes. The former are wrapped up with the latter in such a way that there will often be events at higher levels that influence physical states of affairs, and visa-versa. The naive realist holds that it is the relations of dependence between different causal levels that ensure that there will be a systematic overdetermination of physical events by causes at a variety of different levels; and it is these same relations of dependence that render this overdetermination non-spooky. Insofar as this multiplicity of causal levels is built into the natural world, overdetermination is also built in. The difference between the naturalist and naive realists is that for the former and not the latter, this sort of overdetermination is only innocuous if all high-level causation can

ultimately be understood as physical. The naive realist holds that certain kinds of high-level causation cannot be understood as physical, that there are thus different *sorts* of relations of dependence that a high-level cause can have on the underlying physical activity. In §3.4, I will contrast the sort of relation discussed in §2.1 with an entirely different sort of relation. Either sort is sufficient to remove the threat of violating the following principle:

> *The No Spooky Overdetermination Principle* (NSO): There is no non-coincidental and undesigned overdetermination of events by independent causal histories.

As overdetermination is ordinarily conceived, however, overdetermining causes are thought of as both independent and sufficient for their effects. In rejecting the idea that mental causes are independent of underlying physical processes, I also deny that mental causes suffice for their physical effects; I deny, that is, that nothing but a mental cause is needed for it to have its physical effects. To say they are *dependent* just is to say that they cannot do their work without cooperation from below. And this is, I think, as it should be. It might be that I wanted to catch the bus, so ran, but that's no reason to say that this desire (together with the relevant beliefs) was sufficient for my running. Had the earth opened up in front of me, had my heart stopped beating, I would not have run, and none of the physical events associated with my running would have occurred. Lack of sufficiency in this sense is not a blow to the efficacy of the mind, for the idea that mental states make a difference to what happens *irrespective of our bodily condition* is no part of our ordinary conception of mental causation.[29]

What *is* a part of this ordinary idea is that, had I not wanted to catch the bus (and had I no other reason to run), I would not have run and the physical effects of my desire would not have occurred. It might be thought that this counterfactual implies a violation of CCP, but this is not so. Because of the dependence of the mental on the physical, it will often be the case that if my mental life were different, the underlying physical processes on which my mental life depends would also not have been what they were. It might be, for example, that had I seen a friend across the street, I would not have run to catch the bus. But my physical state would surely also have been different, as my eyes would have undergone a different set of physical changes than they in fact did.[30]

---

[29] On the tenability of the idea of causes that do not determine their effects, see Anscombe's "Causality and Determination" in *The Collected Papers of G.E.M. Anscombe Volume 2* (Minneapolis: University of Minnesota Press, 1981).

[30] For an argument connecting the dependence of the mental on the physical to the truth of this sort of counterfactual, see John Heil and Alfred Mele, "Mental Causation", *American Philosophical Quarterly* vol. 28, no. 1, 1991. Unfortunately, the authors are also committed to the thesis that "the causal clout of a supervenient characteristic [viz., a

Although the rightness of NSO might explain why people are wary of accepting the sort of natural overdetermination of events countenanced by the naive realist, in fact it does nothing to support stronger principles such as SP or WP. There is nothing spooky about the idea that different kinds of causal factors are naturally involved with some physical event's occurrence. It is perfectly compatible with the completeness of its physical causal history.

§3.2 *Mental States as Non-Mechanistic.* The fact that WP is so widely held can be explained by reference to four points that have already been discussed: (1) the thought that a certain kind of systematic overdetermination *would* be spooky, (2) the need to make a distinction between real and bogus high-level properties, (3) the acknowledgment that physical causation is fundamental,[31] (4) the fact that some high-level causal processes *can* be understood as consisting in fundamental causal processes. But though these factors help to explain its appeal, it is important to realize that they do not jointly entail WP. In fact a naive realist is equally able to affirm the truth of (1)–(4).

The project of naturalizing the mind derives from the desire to avoid seeing causal claims involving the mind as violating WP; it is attempted in the spirit of showing that mental properties are not spurious. Such a project derives its urgency from a prior acceptance of WP; it is motivated by the fear that the failure of such a project entails that mental causation is tainted with magic and mysticism. And though I agree with eliminativists on the prospects for understanding the causal powers of the mind mechanistically, I don't think this tells against those powers. I think we ought to reject WP instead.

Herein lies the importance of acknowledging the possibility that naturalists deny, i.e., the possibility of naive realism. While naturalizing the mind looks like the only way to coherently maintain a realistic stance towards mental causation, its program and presuppositions will always earn the benefit of the doubt from philosophers who are unwilling to stomach eliminativism. In this climate, naturalist approaches to the mind enjoy the presumption of correctness that derives from being the only game in town. Once the tenability of non-naturalist realism has been established, however, the burden of proof shifts to those who pursue a naturalist program. For the credibility of these programs now lies entirely in the success of their various specific attempts to locate mental states and their causal powers in the physiological realm. To the extent that these attempts prove unsuccessful we have

---

mental characteristic] resides in whatever realizes that characteristic" (p. 68). As they are committed to the view that mental causation does not amount to anything over and above physical causation, I consider them adherents of WP.

31 Naive realism is compatible with physics being fundamental in the sense suggested by the completeness of physical causal histories and the idea that other causal levels are all dependent on the physical level, in the sense to be discussed in §3.4.

evidence not against realism about the mental, but rather simply naturalism about the mental.

My goal is thus not to *prove* that WP is false. Rather than attack the empirical claim that further research will yield a mechanistic explanation of thoughts, I advance the metaphysical thesis that the absence of such an explanation is compatible with the existence and efficacy of the mental. That is, I attempt to show that naive realism is a viable alternative to naturalist realism.

In sketching the kind of states mental states might be if they are not naturalizable, it will be helpful to have a relatively rich example of a psychological explanation in mind. Consider the following example, based on an event in Buñuel's *Viridiana*.

> Don Jaime has decided to commit suicide. He wants Viridiana not to return to her convent, and comes to believe that by leaving her his estate, he will keep her there. This realization pleases him, causing him to snicker.

The focus on beliefs and desires in the literature makes it easy on the naturalist's imagination. One can try to think of the causally efficacious event behind the coming-to-believe as a network of neurons suddenly becoming excited, a desire making itself felt as a surge of hormones or some such thing. Causation then involves the spread of excitement, the onset of an adrenaline rush or whatever.

But let us ask: What is the psychological causal history of Don Jaime's snickering? What sort of person is capable of snickering at the prospect of undermining, through his own suicide, his niece's plans to return to her convent? There is no short answer to this question. He must be someone whose life appears to himself to be of so little importance that he is not only willing to give it up, but can be in a state of mind that permits snickering only moments before doing so. He must be a very spiteful man, for his glee is not connected to any benefit that he expects to receive. He just likes the idea that her determination to leave his estate will come to naught. The movie indicates that this is connected to an incestuous fascination he's developed with her chastity, a feeling that is somehow connected to his memory of the death of his very young wife. And of course, he only finds his action funny because he is normal in many respects. He knows, for example, how the world works enough to be able to see that Viridiana will be forced to return if he is found dead.

The causal relation between his want, his coming to believe, and his entering a state of snickering is causally mediated by Jaime's self-image, his traumatic memories, his view of the value of life, his strange sexual sensibility, his sense of humor, and his general understanding of the way the world

works, among other things. All of these facts are part of the causal history of his snickering, and we have not really understood the nature of this particular and peculiar instance of causation if we do not understand any of that.

On the face of it, however, there is nothing about any of this that suggests simple mechanical units undergoing interactions that can be characterized in simple mechanical terms. Of course, a smile can be described in mechanical terms, and we know there is a causal story behind the smile, one involving causal transactions between internal objects, viz., neurotransmitters, nerves, muscles, and so forth. But there is nothing about the idea of a sense of humor, or a conception of one's self-worth that suggests that their causal power can be understood in mechanical terms.

When we explain Don Jaime's behavior in terms of mental states, we do not make the sort of commitment about what's going on *inside his body* that we do when we say he is in a biological, chemical or physical state: To put it generally, the existence of a psychological state has no necessary space-involving ontological implications beyond

(1) the fact that there must be a person (among other things, a spatio-temporal continuant) who's in it; and

(2) whatever states of the world are presupposed by the existence of a psychological state with *that* content.[32]

I will use the expression 'minimally space-involving' to refer to this feature of psychological states. Causal explanations of the states of mechanical systems can only be understood in terms of the relationships between the internal material constituents of the system. They are robustly space-involving. But mental states are or in any case might well be minimally space-involving.

The justification for this claim will come in three by-themselves-uninteresting parts. The first part concerns the point that judgments about mental causation are grounded at the level of (common-sense) psychology. The second part involves the idea that the relations that are constitutive of psychological states are rational ones, and the third the idea that the causal role of psychological states is determined by these rational relations. These are all points that have been made before. If there is progress in the offing here, it will come by seeing that these facts do or in any case might well render mental causation unsuitable for mechanistic comprehension. And this need not trouble us, as respect for the importance and informativeness of the

[32] This second condition is meant to account for the causal and constitutive dependence of conceptual content on 'external' factors.

mechanistic sciences does not require that we accept WP. Faced with a tension between mental causation and WP, we should reject WP.

What counts when making a judgment about why someone performed some action (or laughed, or remained silent...) is what else we know about what she believes and desires (or what she finds funny, or what she doesn't know...), what she's said or done in the past, and what she's revealed about her attitude towards what she's said or done in the past. A psychological explanation is judged by how well it rationally coheres with the rest of what is known about her mental life. We know quite a lot about the mechanisms whose proper functioning is necessary for people to go about their business (see §3.4 for more on this), and so we can make good inferences about what must be going on inside their bodies based on facts about what they do. Conversely, we can make good inferences about what, at a certain level of generality, must be going on in their minds from facts about the state of their insides. (E.g., "That's gotta hurt" or "He should be as high as a kite in about fifteen minutes.") But these are not the kinds of inferential connections that make psychological states and events what they are. These inferences reflect what are, in a sense, more peripheral, contingent, conceptual connections.

The identity of psychological states and of actions is constituted by their placement in a network of rational (in the broadest sense) relations, which do not have the connection to spatial concepts characteristic of the relations between elements in mechanistic explanations. In virtue of their contents, psychological states stand in logical relations like incompatibility, material implication, and conceptual necessitation. They also stand in more interesting identity-constituting relations such as being funny/tragic/unfair/untoward in light of other thoughts, actions, and known states of affairs. Bits of our psychological characterizations of people are constitutively dependent on a more or less detailed background understanding of their mental life. We often arrive at psychological understanding by thinking of someone as having a certain character. In doing so, we attribute to them a complex network of interacting dispositions, sensibilities, views on every scale, inclinations, longings, patterns of thought, degrees of consistency in mood and stability of plans, etc. The plasticity of concepts like 'state' allows that we could describe all of these aspects of a person in terms of her psychological states. The rational relations these states stand in are perfectly real, and it is constitutive of our idea of what a mental life is that its elements stand in them.

In ascribing to people states that stand in such relations, we presuppose that the relata enter into causal relations with one another, and that they do so in virtue of the obtaining or not of these relations, i.e., in virtue of their contents. The fact that mental states stand in causal relations is registered in our applying to them a causal vocabulary. It is crucial to our understanding of what longings are, for example, that they can sustain ambitions, be inhibited by feelings of guilt, prompt periods of depression, influence taste in movies,

etc. 'Sustaining', 'inhibiting', 'prompting' and 'influencing' are all causal terms, and their use is governed by the rational relations that define them. Whether my feeling of shame, for example, can be counted as a cause of my forming a particular intention depends on what I am ashamed of and what I've decided to do.

Of course, prior to our having a mechanistic understanding of any domain, there is always room for doubting that one is forthcoming. The history of science is replete with examples of those who thought the mechanistic paradigm would break down in certain domains. Famously, the Vitalists thought that animals have a life-force whose efficacy couldn't be understood in chemical terms. And it turned out that they were wrong. There is no such thing as a life-force. (They also thought there were physical occurrences that could not be accounted for in solely physical terms, which is a claim I deny with CCP.) But this is hardly decisive. And in any case, my discussion of mental states is not offered as showing that we will never understand the mind mechanistically. My claim is that the success of the mechanistic sciences does not entitle us to infer a metaphysical principle that all domains *must* be mechanistically understandable on pain of elimination.

To summarize: Our conception of why people believe, desire, and do what they do is tied to our understanding of the conceptual space of psychological concepts as a causal as well as a rational one, but as having only minimally space-involving ontological commitments. Since it is the rational relations which are both identity-constituting and those in virtue of which psychological states stand in the causal relations they do, there is no reason to think we will *necessarily* get a better view of mental causation by looking at neurophysiological phenomena. The forces at work in causation that can be understood mechanistically have nothing to do with those relations. Nor is there any reason to think this tells against the existence and efficacy of the mind. In making the transition from a psychological causal history to a neurophysiological causal history, we may simply be moving from looking at one kind of causal relation to looking at another.[33] In the next subsection, I will

[33] In "Mental Causation", Crane claims that the problem of mental causation only arises on what he calls the 'homogeneity assumption', according to which "the notion of causation is the same notion applied to the physical and the mental alike"; if mental causation is "utterly weird and *sui generis*", then, according to Crane, there is no conflict between mental causal claims and physical causal claims (p. 219). Insofar as the solution I am here proposing may appear to turn on rejecting that assumption, it might be objected that I have made the problem a trivial one. But in fact, I do not reject the idea that "the notion of cause is the same notion applied to the physical and the mental alike". I here argue rather for the possibility that mental and physical causes are two species of causes. Briefly, mental and physical causes each make a difference to what happens, though they (may) do so in different ways. And the idea that mental and physical causes are causes of different sorts does *not* by itself resolve the problem of mental causation—this is *only* effective once one has defended the possibility of overdetermination by a variety of causes, i.e., difference-making factors.

preempt concerns that this characterization of the mind renders its relationship to the rest of the natural world deeply mysterious.

§3.3 *Naive Realism not Anti-Scientific.* One can hold this view of mental phenomena without denying that there are also neurophysiological processes crucial to any psychological occurrence. The proper functioning of a brain is the 'that without which' of any psychological activity at all. There are specific processes in the brain that bear on specific psychological capacities and tendencies. There is a perfectly good sense of 'explain' in which neurophysiological processes can be said to explain these dependent psychological processes. E.g., our ability to remember new facts is explained by the proper functioning of the hippocampus. Lesions to Wernicke's area sometimes explain deficits in the comprehension of verbal communication. 'To explain' here means to provide the physiological basis.

One could go into great detail in the direction suggested by these explanations. One could also try to give a very general account of their structure. I will do neither of those things here. I only want to make room for the idea that not every species of dependence on lower-level machinery needs to be understood according to the same model. In the previous section, we saw one model, according to which a lower-level process (mechanistically) implements a higher-level process. In describing mental processes, we may need a different model, one according to which a lower-level process (non-mechanistically) subserves the higher-level process. It may be that there is nothing more to this relation of subservience than the idea that the proper functioning of the subserving machinery is a necessary condition for the high-level process, with room left for the possibility of correlations between specific types of phenomena at the two levels. Of course, our conception of what it is for our biological machinery to be functioning properly is parasitic on our conception of what forms human life normally takes. Hence the relationship is normative at both ends—for a human brain to be in good working order is for it to make possible a normal conscious life. (Here is how I sever the tie between the 'perfectly sensible thought' of §2.1 and WP.)

The important point here is that explanations of psychological processes in terms of our biological machinery are not psychological explanations. And part of the point of this paper is that there is no need to think of these as competing with or correcting or replacing psychological explanations. The causal power of mental states may not be identifiable with the causal powers of neurophysiological states, and this is no cause for alarm. They are explanatory in different ways; they reflect different aspects of the causal structure of the world.

We can make the same point in terms of causal histories. Every psychological state of affairs has a physiological causal history. My claim that mental causation might not be describable in mechanistic terms is in no way

at odds with that fact. I only insist that we recognize and respect the difference between causal histories at the level of psychology and those at mechanistic causal levels.

Naive realism is thus also compatible with various kinds of supervenience relations between the mental and physical realms. It is compatible, for example, with 'weak' or 'global' supervenience—the idea that physically identical worlds are identical in every respect. I have not focused on the concept of supervenience in this paper, in part because there is very little agreement on how much it can be strengthened, and in part because there's even less agreement about what the significance of a strengthened version's obtaining or not obtaining would be. But also, I think the kind of relation that naturalists (falsely) think must hold between higher and lower levels of explanation is captured just as well by the concept of an implementing mechanism as by the idea of a strong-supervenience-base.

Conclusion

Consider these three possible reactions to the apparent tension between the manifest and scientific images of human beings: eliminativism, naturalist realism, and naive realism. The debate over mental causation can be understood as fueled by concerns about overdetermination—the idea that the different kinds of causal claims associated with the two images cannot both be accommodated. In the case of the most radical expression of this concern, The Strong Principle of Non-Overdetermination, the tension between physics and common sense can be resolved only by abandoning common-sense altogether: eliminativism. The Weak Principle of Non-Overdetermination leaves matters this way: If you think that the mind can be mechanistically explained, you are a naturalist realist. If you think it can't, then you must be an epiphenomenalist or an eliminativist. But in failing to see WP as something that needs to be defended, these naturalists have artificially limited the field of legitimate possibilities.

Among the three groups there are various points of agreement and disagreement. Naturalist realists and eliminativists agree that the mind must be describable in the language of natural science if it is to be accepted into our ontology, but disagree over whether this is possible. Naive realists and eliminativists agree that no scientific account of the mind is possible, but disagree about what the significance of this impossibility is. Naturalist and naive realists agree on this: "Mental causes are not in competition with physical causes. They are merely causes at a higher level than those mentioned by physics, as are causes proper to economics, mechanical engineering, geology, biology, chemistry, and so forth. Mental causation is associated with distinct causal-explanatory principles, but the same systems can be governed by more than one set of such principles." In the mouths of naive and naturalist realists, however, this response will mean very different

things. For the naturalist, the legitimacy of any non-fundamental science (its naturalizability) depends on being able to see the causal powers of the relevant systems as mechanistically implemented by their physical constituents. The naive realist rejects the idea that this constraint can be applied universally. Naive realism is the view that the manifest image of human beings can peacefully coexist with the scientific image, and can do so without the distortions wrought by 'naturalization'. The best defense of the tenability of this view involves showing how specific attempts to bring the two views into conflict fail, and that is what I've tried to do here.

In this paper, I have not set out to prove that naive realism is true, that the mind *can't* be naturalized. I have only sought to show that naive realism is *possible*. This might appear to be an uninteresting claim. To see why it is not, consider the following distinction between what we can call *empirical naturalism* and *metaphysical naturalism*. According to empirical naturalism all causation can in fact be understood as physical causation. According to metaphysical naturalism, all causation *must* be understandable as physical causation. For the empirical naturalist, it is, until settled, an open question whether mental causation can be understood physicalistically. Her hypothesis is that it can be so understood, but she would draw no eliminativist consequences from the falsity of that hypothesis. The metaphysical naturalist, on the other hand, thinks that it is not an open question whether mental causation can be understood physicalistically. Either there is mental causation, and therefore it can be so understood, or there is no mental causation.

If naive realism were true, this would show that empirical naturalism is false. That naive realism is possible shows that metaphysical naturalism is false. And it is the possibility-claim that I have been concerned to establish here. Significantly, the falsity of metaphysical naturalism enables us to hear arguments against empirical naturalism in the right way. Until metaphysical naturalism is rejected, arguments against the hypothesis of the empirical naturalist (e.g., from externalist considerations, from considerations involving the distinctive normative character of the mental, from holism about the mental, from facts about qualia or consciousness) will always be open to eliminativist interpretation. And that is why it is crucial first to recognize the possibility of naive realism before considering the merits of criticisms of empirical naturalism. Furthermore, once one recognizes this possibility, the urgency of the empirical naturalist's project is greatly diminished; for the tenability of the manifest image of human beings no longer hangs in the balance.[34]

[34] I am very grateful to John Haugeland, James Conant, John McDowell, Ram Neta, Doug Lavin, Doug Patterson, Carl Craver, and an anonymous referee at this journal for helpful suggestions and criticisms.