Mental Causation

Stephen Yablo

The Philosophical Review, Volume 101, Issue 2 (Apr., 1992), 245-280.

Stable URL:
http://links.jstor.org/sici?sici=0031-8108%28199204%29101%3A2%3C245%3AMC%3E2.0.CO%3B2-8

The Philosophical Review, Vol. 101, No. 2 (April 1992)

Mental Causation*

Stephen Yablo

1.

Writing to Descartes in 1643, Princess Elisabeth of Bohemia requests an explanation of "how man's soul, being only a thinking substance, can determine animal spirits so as to cause voluntary actions."[1] Agreeing that "the question which your Highness raises [is] one which can most reasonably be asked," Descartes launches with his reply a grand tradition of dualist apologetics about mind-body causation that has disappointed ever since. Apologetics are in order because, as Descartes appreciates, his conception of mental and physical as metaphysically separate invites the question, "how, in that case, does the one manage to affect the other?"; and because having invited the question, he seems unable to answer it. Much as the Cartesian epistemology breeds skepticism, then, the metaphysics breeds epiphenomenalism: the theory that our mental lives exercise no causal influence whatever over the progress of physical events.

That was the price Descartes paid for his dualism, someone

*Thanks to Louise Antony, Simon Blackburn, Paul Boghossian, Donald Davidson, Graeme Forbes, Sally Haslanger, Jaegwon Kim, Vann McGee, Sarah Patterson, Gideon Rosen, Larry Sklar, William Taschek, David Velleman, Ken Walton, Catherine Wright, Crispin Wright, and two anonymous readers for reactions and advice. Versions of the paper were read at Chicago Circle, the University of North Carolina at Chapel Hill, and the University of Western Ontario; discussions there were extremely useful and I'm grateful to all who took part. Research was supported by the National Endowment for the Humanities and the Social Sciences and Humanities Research Council of Canada.

[1] *The Essential Descartes*, ed. M. Wilson (New York: New American Library, 1969), 373. In the "Fifth Objections," Gassendi puts a similar question: "How can there be effort directed against anything, or motion set up in it, unless there is mutual contact between what moves and what is moved? And how can there be contact without a body . . . ?" (*The Philosophical Writings of Descartes*, vol. 2, ed. J. Cottingham, R. Stoothoff, and D. Murdoch [Cambridge: Cambridge University Press, 1984], 236ff.).

might say. Why should epiphenomenalism concern anyone today? Part of the answer is that dualism is not dead, only evolved. Immaterial minds are gone, it is true, but mental *phenomena* (facts, properties, events) remain. And although the latter are admitted to be physically *realized*, and physically *necessitated*, their literal numerical *identity* with their physical bases is roundly denied.[2]

Surely, though, it is hard to imagine a dualism more congenial to mental causation than this! So it would seem. But epiphenomenalism has been evolving too; and in its latest and boldest manifestation, this is all the dualism it asks for. As a result we find ourselves in a somewhat paradoxical situation. Just when the conditions for accommodating mental causation have become little short of ideal, epiphenomenalist anxiety rages higher than ever. Nor is this a pretended anxiety, put on for dialectical purposes but posing no genuine danger to established views. Some say we must simply make our peace with the fact that "the mental does not enjoy its own independent causal powers."[3] Others would renounce (distinctively) mental phenomena altogether, rather than see them causally disabled.[4] Radical as these proposals are, they are backed by a straightforward line of reasoning.

"How can mental phenomena affect what happens physically? Every physical outcome is causally assured already by preexisting physical circumstances; its mental antecedents are therefore left with nothing further to contribute." This is the *exclusion argument*

---

[2]In case it seems odd to describe the picture just outlined as dualist, bear in mind that all I mean by the term is that mental and physical phenomena are, contrary to the identity theory, *distinct*, and contrary to eliminativism, *existents*. That this much dualism is acceptable even to many materialists is in a way the point: having broken with dualism's Cartesian version over its vulnerability to epiphenomenalism, they find to their horror that epiphenomenalism lives equally happily on the lesser dualism latent in their own view.

[3]Kim, "Supervenience and Supervenient Causation," *Southern Journal of Philosophy*, supp. vol. 22 (1983): 54. Kim does allow the mental a role in what he calls *epiphenomenal* causal relations, and he says that macrophysical causation is epiphenomenal in the same sense. My position is that neither sort of causation is epiphenomenal in any interesting sense.

[4]This is particularly clear in Schiffer, who rejects mental *properties* on the ground that they would be causally superfluous, and makes mental *events* a subspecies of physical events on the theory that they would *otherwise* be causally superfluous (*Remnants of Meaning* [Cambridge: MIT Press, 1989], chap. 6).

for epiphenomenalism. Here is the argument as it applies to mental events; for the version which applies to properties, replace 'event x' with 'property X':[5]

(1) If an event x is causally sufficient for an event y, then no event x^* distinct from x is causally relevant to y (*exclusion*).[6]
(2) For every physical event y, some physical event x is causally sufficient for y (*physical determinism*).[7]

---

[5]So 'x' and 'x^*' become 'X' and 'X^*', and where either is prefixed by 'event', this becomes 'property'; 'event y' and 'event z' are unaffected. Although causes and effects are events, properties as well as events can be causally relevant or sufficient. I try to remain neutral about what exactly causal sufficiency and relevance amount to (e.g., causal sufficiency could be sufficiency-in-the-circumstances, or it could be absolute). Versions of the exclusion argument are found in H. Feigl, "Mind-Body, Not a Pseudo-Problem," in *The Mind-Brain Identity Theory*, ed. C. V. Borst (New York: St. Martin's Press, 1970), 33–41; N. Malcolm, "The Conceivability of Mechanism," in *Free Will*, ed. G. Watson (Oxford: Oxford University Press, 1982), 127–49; A. Goldman, "The Compatibility of Mechanism and Purpose," *Philosophical Review* 78 (1969): 468–82; K. Campbell, *Body and Mind* (New York: Macmillan, 1970); J. Kim, "Causality, Identity, and Supervenience in the Mind-Body Problem," *Midwest Studies in Philosophy* 4 (1979): 31–50, and "Mechanism, Purpose, and Explanatory Exclusion," *Philosophical Perspectives* 3 (1989): 77–108; E. Sosa, "Mind-Body Interaction and Supervenient Causation," *Midwest Studies in Philosophy* 9 (1984): 271–81; T. Honderich, *Mind and Brain: A Theory of Determinism* (Oxford: Oxford University Press, 1988); and C. Macdonald and G. Macdonald, "Mental Causation and Explanation of Action," in *Mind, Causation, and Action*, ed. L. Stevenson, R. Squires, and J. Haldane (Oxford: Basil Blackwell, 1986), 35–48. Objections similar in spirit to the exclusion argument are sometimes raised against the causal claims of other phenomena apparently unneeded in fundamental physical explanation (e.g., macroscopic and color phenomena). This paper offers a potentially general strategy of response.

[6]Some authors use a slightly weaker premise: if x is causally sufficient for y, then unless y is causally overdetermined, every distinct event x^* is causally irrelevant (see note 53).

[7](2) could obviously be questioned, but I take it that physical determinism isn't the issue. For one thing, the conviction that mind makes a causal difference is not beholden to the contemporary opinion that determinism is false, and would remain if that opinion were reversed. Second, nothing essential is lost if 'x is causally sufficient for y' is replaced throughout by 'x determines y's objective probability'. So unless the argument can be faulted on other grounds, mental causation is problematic under indeterminism too.

(3) For every physical event x and mental event x^*, x is distinct from x^* (*dualism*).
(4) So: for every physical event y, no mental event x^* is causally relevant to y (*epiphenomenalism*).

This is bad enough—as Malcolm says in "The Conceivability of Mechanism," it means that no one ever speaks or acts—but a simple extension of the argument promises to deprive mental phenomena of all causal influence whatsoever. Every event z of whatever type is metaphysically necessitated by some underlying physical event y, whose causally sufficient physical antecedents are presumably sufficient for z as well. But then by the exclusion principle, z's mental antecedents are irrelevant to its occurrence. So, mental phenomena are *absolutely* causally inert. And now it is not only speech and action that are chimerical but also thinking.

Note well that the exclusion argument raises *two* problems for mental causation, one about mental particulars (events), the other about mental properties.[8] Strangely, philosophers have tended to treat these problems in isolation and to favor different strategies of solution.[9] In Malcolm's original presentation, he emphasizes problem one. Given a neurophysiological theory rich enough to

> provide sufficient causal conditions for every human movement, . . . there would be no cases at all in which [the] movement would not have occurred if the person had not had [the] desire or intention . . . [thus] desires and intentions would not be causes of human movements.[10]

Here the mystery is how mental *events*, desires for example, can be making a causal difference when their unsupplemented neurophysiological underpinnings are already sufficient to the task at hand. To reply with the majority that mental events just *are* certain

[8]C. D. Broad was perhaps the first to emphasize epiphenomenalism's double-sidedness: "[it] asserts . . . that mental events either (a) do not function at all as cause-factors; or (b) that, if they do, they do so in virtue of their physiological characteristics and not in virtue of their mental characteristics" (*Mind and its Place in Nature* [London: Routledge and Kegan Paul, 1925], 473).

[9]Kim, "Epiphenomenal and Supervenient Causation" (*Midwest Studies in Philosophy* 9 [1984]: 257–70) is an important exception.

[10]Malcolm, "The Conceivability of Mechanism," 136.

physical events, whose causal powers they therefore share,[11] only relocates the problem from the particulars to their universal features:

> the being of a desire by my desire has no causal relevance to my extending my hand . . . if the event that is in fact my desire had not been my desire but had remained a neurological event of a certain sort, then it would have caused my extending my hand just the same.[12]

Mental events are effective, maybe, but not by way of their mental *properties*; any causal role that the latter might have hoped to play is occupied already by their physical rivals.[13] Although someone *could*, following the line above, attempt to *identify* mental properties with (certain) physical properties, say, being a desire with instantiating such and such a neurophysiological type, this approach is

[11]See Feigl, "Mind-Body, Not a Pseudo-Problem," 36ff.; J. Smart, "Sensations and Brain Processes," in Borst, *The Mind-Brain Identity Theory*, 54, 65–66; and Davidson, "Mental Events," in *Essays on Actions and Events* (Oxford: Oxford University Press, 1980), 207–24. Note that Davidson advances the token identity theory in response to a slightly different problem. His aim is to reconcile the following assumptions: singular causal claims need always to be backed by strict causal laws; strict laws are physical laws; every event subsumable under a physical law is a physical event; and mental events are efficacious.

[12]Sosa, "Mind-Body Interaction," 278.

[13]Again, this needs to be distinguished from a somewhat different worry directed primarily at Davidson's anomalous monism: singular causal claims need always to be backed by strict causal laws; *x*'s causally relevant properties *vis-à-vis y* are those figuring in the antecedent of some such backing law; strict causal laws never involve mental properties; so *x*'s mental properties are causally irrelevant. For discussion, see Stoutland, "Oblique Causation and Reasons for Action," *Synthese* 43 (1980): 351–67; Honderich, "The Argument for Anomalous Monism," *Analysis* 42 (1982): 59–64; Sosa, "Mind-Body Interaction"; Loewer and Lepore, "Mind Matters," *Journal of Philosophy* 84 (1987): 630–42; Fodor, "Making Mind Matter More," *Philosophical Topics* 17 (1989): 59–79; Loewer and Lepore, "More on Making Mind Matter More," same volume: 175–91; Cynthia Macdonald and Graham Macdonald, "Mental Causation and Explanation of Action"; and Brian McLaughlin, "Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical," *Philosophical Perspectives* 3 (1989): 109–35 (some of these papers discuss the exclusion objection also). Note that the exclusion objection, the subject of the present paper, assumes nothing about the role of laws in causation or in the characterization of causally relevant properties.

now discredited, because of the well-known multiple realizability objection.[14] Properties are identical only if each necessitates the other; but any physical property specific enough to necessitate a mental property is inevitably *so* specific that the converse necessitation fails. Since (as I'll maintain) the objection applies, *mutatis mutandis*, to mental *particulars*, the identity response is unworkable in either case.[15]

So I find no fault with dualism, or with the associated picture of mental phenomena as necessitated by physical phenomena which they are possible without. Rather than objecting, in fact, to the asymmetric necessitation picture, I propose to go it one better. Traditionally, the paradigm of one-way necessitation was the relation of *determinate* to *determinable* (sections 2 and 5). What if mental phenomena are determinables of physical phenomena in something like the traditional sense (sections 3 and 6)? Then since a determinate cannot preempt its own determinable, mental events and properties lose nothing in causal relevance to their physical bases (sections 4 and 7).[16] If anything, it is the other way around. Overladen as they frequently are with physical details far beyond the effect's causal requirements, it is the *physical* phenomena which are liable to disqualification on grounds of superfluity (section 8).

2.

Before asking what determinates and determinables might be, consider the "easier" question of when properties are identical. Probably no one would quarrel with

---

[14]See, for example, H. Putnam, "The Nature of Mental States," and N. Block and J. Fodor, "What Psychological States are Not," both in *Readings in Philosophy of Psychology*, vol. 1, ed. Block (Cambridge: Cambridge University Press, 1980).

[15]This is hardly a cause for regret. Identifying mental phenomena with physical phenomena, we saddle the former with the causal properties of the latter; but common sense sees mental phenomena as possessed of *distinctive* causal properties (see sections 8 and 9).

[16]About mental and physical *properties*, the Macdonalds ("Mental Causation and Explanation of Action") reach a similar conclusion; however, their argument depends on treating mental *events* as identical to, rather than determinables of, physical events (see note 32 for the problems this causes).

(I) *P* is identical to *Q* iff: for a thing to be *P* is for it to be *Q*,

on at least some interpretation. But, apart from its possible circularity, **(I)** explains one obscurity with another; and it has become customary to seek relief from both complaints in the modal idiom. That idiom permits no sufficient condition for property identity, unfortunately; so something is sacrificed. But we're repaid with the necessary condition that

(I) $P = Q$ only if: necessarily, for all x, x has P iff x has Q.[17]

Properties are identical, in other words, only if it is impossible for a thing to possess either without possessing the other.

Among (I)'s attractions is that we *know* it is true since it follows from Leibniz's Law, the indiscernibility of identicals. Or better: it follows if the modality is read as *metaphysical*. Whether because they conflated conceptual with metaphysical necessity, or because they construed the properties themselves as concepts, philosophers *used* to think that properties were the same only if it was *conceptually* or *a priori*[18] true that their instances could not differ.[19] (Thus they felt justified in arguing from purely conceptual considerations to a distinction between, say, being salt and being sodium chloride.) This stronger condition can of course claim no support from Leibniz's Law.[20] But that isn't what led to its rejection: it was rejected because it proved unable to cope with the discovery of identical

[17]Treating necessary coextensiveness as also *sufficient* for property identity would lead to various unwanted results, for instance, that there is only one universally necessary property.

[18]I lump these two together not out of conviction but just as an expedient.

[19]This, the condition (I_1) that properties are identical only if their *necessary* coextensiveness is conceptually guaranteed, entails (I) trivially; (I) does not entail (I_1) conversely because some necessary coextensiveness claims are not *a priori* knowable, for example, that necessarily, the extension of identity-with-Hesperus is the same as that of identity-with-Phosphorus. Note the contrast between (I_1) and the weaker condition (I_2) that $P = Q$ only if their *actual* coextensiveness is knowable *a priori*. (I_1) and (I_2) fail for essentially similar reasons (see note 21), but it is (I_1) that I have in mind in the text.

[20]Reason: 'it is *a priori* that . . . ', like 'Jones believes that . . . ', generates an opaque context.

properties, such as the ones just mentioned, whose necessary coextensiveness was knowable only *a posteriori*.[21] So the mutual conceptual necessitation requirement is now defunct; its metaphysical kernel (I), although insufficient for property identity, is the only game in town.

According to a still reputable traditional doctrine, some properties stand to others as *determinate* to *determinable*—for example, *crimson* is a determinate of the determinable *red*, *red* is a determinate of *colored*, and so on.[22] Since the distinction is relative, one does better to speak of a determination *relation* among properties, where

(Δ) P determines Q iff: for a thing to be P is for it to be Q, not *simpliciter*, but in a specific way.

Except for the 'not *simpliciter* . . .', (Δ) would describe identity; and like identity, determination as traditionally understood involves conceptual and metaphysical elements jumbled confusingly together. Metaphysically, the central idea is that

(Δ) P determines Q ($P > Q$) only if:
(i) necessarily, for all x, if x has P then x has Q; and
(ii) possibly, for some x, x has Q but lacks P.

Not always distinguished from this is a requirement of asymmetric conceptual entailment: there is no conceptual difficulty about a world in which some Q lacks P, but the converse scenario is excludable on *a priori* grounds.

Now, just as the discovery of *a posteriori* necessities upset the traditional presumption of a conceptual equivalence condition on property *identity*, it also makes trouble for the conceptual entailment condition on *determination*. Take the property of being at

[21] Kripke, *Naming and Necessity* (Cambridge: Harvard University Press, 1980). Likewise, the weaker condition (I_2) cited in note 19 was overturned by the discovery of identical properties whose *actual* coextensiveness was not knowable *a priori* (e.g., identity-with-Hesperus and identity-with-Phosphorus).

[22] Two classic discussions are W. E. Johnson, *Logic* (New York: Dover, 1964), vol. 1, chap. 11, and Arthur Prior, "Determinables, Determinates and Determinants (I, II)," *Mind* 58 (1949): 1–20, 178–94.

temperature 95° C, and some highly specific micromechanical property K chosen so that necessarily whatever has K has the temperature property, though not conversely. Since Ks which are warmer than 95° C cannot be ruled out on *a priori* grounds alone, traditional determination fails. Yet the relevance of this to the properties' strictly *metaphysical* relations is obscure; and since it is only the metaphysics that matters to causation, we should discount the traditional doctrine's conceptual component and reconceive determination in wholly metaphysical terms.[23] What justifies the continued use of the word 'determine' is that **(Δ)** holds essentially as before. To be in the micromechanical condition of this steaming tea, for instance, is to be at temperature 95° C *in a certain micromechanical way.*

3.

As I write, I am in a certain overall physical condition, and I am also thinking; presumably the one fact about me has quite a lot to do with the other. Suppose the pertinent aspects of my physical condition to be encoded in some physical property P. Could it be that P is a *determinate* of thinking? Barring some unsuspected conceptual entailment from physics to thought, the full-scale traditional doctrine answers in the negative. On the other hand, traditional determination incorporates elements visibly irrelevant to how the properties are related in themselves; so the interesting question is whether P determines thinking in the *metaphysical* sense.[24] I say that it does. And I hold further that there is this sort

[23]So P determines Q just in case the traditional relation's first, metaphysical component is in place, where this consists primarily in the fact that P necessitates Q asymmetrically. Probably it goes too far to identify determination with asymmetric necessitation outright; otherwise, for example, conjunctive properties determine their conjuncts and universally impossible properties are all-determining. For dialectical reasons, I try to remain as neutral as I can about where determination leaves off and "mere" asymmetric necessitation begins (Prior, "Determinables, Determinates and Determinants," reviews some of the fascinating history of this problem).

[24]"But if there is no conceptual entailment from P to thinking, then unthinking Ps are conceivable, and to that extent possible; thus P doesn't determine thinking in the metaphysical sense either." I grant that the conceivability of a proposition ϕ is *prima facie* evidence of its possibility. But

of physical determination whenever a mental property is exemplified.

Such a view is in fact implicit in the reigning orthodoxy about mind-body relations, namely, that the mental is *supervenient* on, but *multiply realizable* in, the physical.[25] Because neither thesis concerns determination directly, the point is easily missed that in combination their effect is to portray mental properties as determinables of their physical realizations. Take supervenience first, the claim that a thing's mental properties are fixed by how it is physically:

(S) Necessarily, for every x and every mental property M of x, x has some physical property P such that necessarily all Ps are Ms.[26]

---

this *prima facie* evidence is defeated if there is not improbably a proposition ψ such that (a) ϕ is true, (b) if ψ is true, then ϕ is impossible, and (c) ϕ is conceivable only because one was unaware of (a) and/or (b). The ancients, for instance, were able to conceive Hesperus as existing without Phosphorus only because they were unaware of their identity; and if I find it conceivable that something should be in the micromechanical condition of this steaming tea but with a different temperature, that is for ignorance of the temperature's microphysical explanation. But I take it that there may also be an explanation of how thinking arises out of neurophysiology, such that if I knew it, then I would find it *in*conceivable, and consider it impossible, that something should be P without thinking. What's more, the prospect of such an explanation makes the hypothesis of an unthinking P only dubiously conceivable *today*. So the complaint is questionable on two counts. First, from a proposition's conceptual coherence, from the fact that its denial is not conceptually false, its conceivability does not follow—witness the Hesperus/Phosphorus example. Even where conceptual difficulties are absent, conceivability can be inhibited by the knowledge or suspicion of a defeater; and this is how it is, for many of us, with the proposition that there could be Ps that did not think. Second, any conceivability intuition I *might* muster in this area I regard as unreliable, because liable to defeat by the progress of science. (For the (a), (b), (c) model of modal error, see Yablo, "The Real Distinction Between Mind and Body," *Canadian Journal of Philosophy*, supp. vol. 16 [1990]: 149–201, and "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research* [1993].)

[25]"All but explicit" would not be much of an exaggeration; determination lies so near the surface and so neatly organizes received opinion that one wonders why it is not already a standard theme.

[26]This is Kim's "strong supervenience" ("Concepts of Supervenience," *Philosophy and Phenomenological Research* 45 [1984]: 153–76). Perhaps not everyone accepts supervenience in quite this strong a form; perhaps I don't myself (Yablo, "The Real Distinction Between Mind and Body"). Yet for two reasons I have thought it better to formulate the thesis as in the

Now, thinking is a mental property, and I possess it. By supervenience, then, I have a physical property *P* given which thinking is metaphysically guaranteed. Of course, *P* can be considered a determination of thinking only if it is possible to think *without P*, which is to say otherwise than by way of the physical property that *does* realize my thinking; and this is where the official story's second element comes in.

When philosophers abandoned the hope of finding for every mental property an identical physical property, the reason was that mental properties seemed intuitively to be multiply realizable in the physical.[27] However, some care should be taken about what this means. Is the claim that for *any* pair of properties, one mental and the other physical, something could have the first without the second? Really, this is stronger than intended, or needed. Imagine someone who holds that necessarily every thinker is spatially extended. Surely such a person could accept multiple realization, intuitively understood, without falling into inconsistency; yet since the necessitation of extension by thinking is the necessitation of a physical property by a mental one, her view actually runs contrary to multiple realization as just explained. Provided that they are suitably unspecific, then, physical properties *can* be necessitated by mental properties compatibly with multiple realization—which suggests as the thesis's proper formulation that *M* necessitates no physical *P* that is *specific enough to necessitate M in return*:

(M) Necessarily, for every mental property *M*, and every physical property *P* which necessitates *M*, possibly something possesses *M* but not *P*.[28]

---

text: (i) strong supervenience is seen nowadays not as the *answer* to epiphenomenalism but rather as the context in which the problem as currently discussed arises (avoiding epiphenomenalism may indeed have been part of the original impulse behind (S), but that is what makes its reappearance *under* (S) all the more troubling); (ii) it focuses the essential line of thought to work within relatively strong assumptions. How much supervenience the approach really needs, and whether that much is plausible, are questions for another paper. For now I just state my hope of getting by with a form of supervenience that allows for the possibility of nonphysical thinkers (see note 47).

[27]See Putnam, "The Nature of Mental States," and Block and Fodor, "What Psychological States are Not."

[28]"Now you contradict yourself, for (M) is incompatible with superve-

For purposes of refuting the identity theory, note, (M) is all that's required. If M were P, then P would necessitate it. But then by (M), it could not necessitate P in return, contrary to their assumed identity.

Together, (M) and (S) make it a matter of necessity that something has a mental property iff it has a physical property by which that mental property is asymmetrically necessitated. But this is extremely suggestive, for with 'determines' substituted for 'asymmetrically necessitates', it becomes

(D) Necessarily, something has a mental property iff it has also a physical determination of that mental property;

and (D) is an instance of the standard equation for determinables and determinates generally, namely, that something has a determinable property iff it has some determinate falling thereunder. This calls out for explanation, and the one that comes first to mind is that mental/physical relations are a species of determinable/determinate relations. "Can you really be saying that mental properties stand to their physical realizations in the relation that rectangularity bears to squareness, or that colors bear to their shades?"[29] Yes. At least that is my conjecture, to be evaluated like

---

nience. Let $\vee P_i$ be the disjunction of all M-necessitating physical properties (alternatively, the second-order property of possessing some P_i or other); then (S) entails that M and $\vee P_i$ necessitate each other, contrary to (M)'s claim that physical properties necessitate mental properties only asymmetrically." To respond by denying the reality of disjunctive properties, on the principle that co-possessors of *real* properties are thereby similar, forgets that the $\vee P_i$s *are* similar in that they have M in common. However, a related point still holds good: sharing of *physical* properties should make for *physical* similarity, and unless the multiple realizability thesis can be faulted on other grounds, the $\vee P_i$s are only mentally alike. (The tendency to think of the physical properties as closed under disjunction may owe something to a confusion of wide- and narrow-scope readings of 'x exemplifies a P_i'. What is true is that for each P_i, whether x possesses *it* is a physical question; this does not make it a physical question whether x has some P_i or other.)

[29]"There is a crucial difference: My mental properties *result* from my physical condition, but in no sense does a thing's redness result from its being scarlet." Actually this raises a subtle interpretive question about supervenience. On the *emergence* interpretation, a thing's physical properties are metaphysically prior to its mental properties and bring them into being. To caricature emergentism just slightly, supervenience is a kind of

any other by the evidence for it and by its theoretical fruitfulness. The evidence is as just described; its consequences for mental causation are considered next.

4.

Imagine a pigeon, Sophie, conditioned to peck at red to the exclusion of other colors; a red triangle is presented, and Sophie pecks. Most people would say that the redness was causally relevant to her pecking, even that this was a paradigm case of causal relevance. But wait! I forgot to mention that the triangle in question was a specific shade of red: scarlet. Assuming that the scarlet was causally sufficient for the pecking, we can conclude by the exclusion principle that every *other* property was irrelevant. Apparently, then, the redness, although it looked to be *precisely* what Sophie was responding to, makes in reality no causal contribution whatever. Another example concerns properties of events. Suppose that the structures in a certain region, though built to withstand lesser earthquakes, are in the event of a *violent* earthquake—one registering over five on the Richter scale—causally guaranteed to fall. When one unexpectedly hits, and the buildings collapse, one property of the earthquake that seems relevant to their doing so is that it was violent. Or so you might think, until I add that this particular

---

"supercausation" which improves on the original in that supercauses act *immediately* and metaphysically *guarantee* their supereffects (the supervenience/causation analogy is common; see, e.g., Kim, "Concepts of Supervenience"). Another view is that the supervening mental properties are *immanent* in their physical bases; rather than giving rise to thought by some obscure metaphysical motion, certain material conditions are inherently conditions of thinking. Now, as the objector suggests, immanentism is clearly correct in standard cases of conceptual entailment, for example, scarlet and red, squareness and rectangularity. Surely, though, this ought to make us suspicious about emergentism as an interpretation of the other cases—for how can the properties' conceptual relations bear on the metaphysical character of the supervenience? That the emergentist thinks they do hints at an unconscious appeal to the neo-Humean prejudice that regularities divide into the conceptual and the causal, or causal-like. But the dilemma is unreal: 'whatever is in the micromechanical condition of this tea is at temperature 95° C' fits into neither category, and I see no reason to treat 'whatever is in the physical condition of this person is thinking' differently. On the immanence model, of course, the alleged disanalogy with colors and their shades evaporates.

earthquake was *barely* violent (its Richter magnitude was over five but less than six). What with the earthquake's *bare* violence being *already* causally sufficient for the effect, that it was *violent* made no causal difference.

Surprising results! To the untrained eye, the redness and the violence are *paradigm cases* of causal relevance, but only a little philosophy is needed to set matters straight. Now, though, one begins to wonder: if even paradigm cases of causal relevance fail the exclusion test, what passes it? Not much, it turns out. Almost whenever a property Q is *prima facie* relevant to an effect, a causally sufficient determination Q' of Q can be found to expose it as irrelevant after all.[30] Applying the argument to Q', Q'', etc. in turn, it appears that only ultimate determinates—properties unamenable to further determination—can hope to retain their causal standing.

Or, on second thought, maybe not them either. Not everything about a cause contributes to its effect; and even where a property does contribute, it need not do so in all its aspects. From the examples it is clear that such irrelevancies do indeed creep in, as we pass from determinable to determinate (e.g., registering less than six); and if the determination process is continued *ad finem*, they may be expected to accumulate significantly. So any ultimate determinate seems likely to incorporate causally extraneous detail. But then, abstracting some or all of this detail away should leave a determinable which, since it falls short of the original only in irrelevant respects, is no less sufficient for the effect.[31] By the exclusion

[30]Depending on what exactly the exclusion principle demands in the way of causal sufficiency, Q' might be a determination of Q only in a fairly relaxed sense (see notes 5 and 23). Those uncomfortable about this should remember the dialectical context: we are trying to show that the assumption needed to disempower mental properties—namely, that determinates are causally competitive with their determinables—would, if true, disempower virtually *all* properties. But if they are causally competitive on a *strict* reading of the determination relation, then when it is *loosely* construed they should be competitive also; and the argument in the text, with determination read the second way, shows that this results in a basically unmeetable standard of causal relevance.

[31]Although it contributed nothing to the earthquake's destructiveness that it registered under Richter six, a determinate of its violence that omitted this would *ipso facto* not be ultimate. Hence the ultimate determinate, whatever exactly it may be, sets a causally idle upper bound on the earthquake's violence; abstracting this upper bound away, we arrive at a

principle, this robs even ultimate determinates of their causal powers. And now it begins to look as though no property ever makes any causal difference.

At least as it applies to properties, then, the exclusion principle is badly overdrawn. Not that there is nothing right about it. In *some* sense of 'separate', it stands to reason, separate properties *are* causal rivals as the principle says. Then what if someone identifies the appropriate notion of separateness and reformulates the exclusion principle accordingly? Suppose it done. Even without hearing the details, we *know* that the corrected principle does not apply to determinates and their determinables—for we know that they are not causal rivals. This kind of position is of course familiar from other contexts. Take for example the claim that a space completely filled by one object can contain no other. Then are even the object's *parts* crowded out? No. In this competition wholes and parts are not on opposing teams; hence any principle that puts them there needs rethinking. Likewise any credible reconstruction of the exclusion principle must respect the truism that determinates do not contend with their determinables for causal influence.[32]

---

determinable still sufficient for the buildings' collapse. (Again, in some cases, this might be a determinable of the ultimate determinate only in a fairly relaxed sense—but see the previous note.)

[32]This is the Macdonalds' view also, but I question their rationale. Sometimes they seem to be arguing as follows: properties derive their causal powers from their instances; if one property determines another, an instance of the first is an instance of the second; so whenever a determinate is efficacious, its determinables are too. However, the conclusion is much too strong. Imagine a glass which shatters if Ella sings at 70 decibels or more. Tonight, as it happens, she sang at 80 db, with predictable results. Although it was relevant to the glass's shattering that the volume was *80* db, it contributed nothing that it was *under 90* db. Therefore, an efficacious determinate can have an irrelevant determinable. Another reading of the Macdonalds' position might be that the determinate's instances are instances of the determinable only *sometimes*, and that it is only in *these* cases that the determinable is efficacious if the determinate is. But notice what this requires: Ella's singing at 80 db is *identical* to her singing at over 70 db, but *distinct* from her singing at under 90 db. Apart from its intrinsic implausibility, such a view is untenable for logical reasons. P and its determinable Q are efficacious not absolutely, but only relative to some specified effect; whether their instantiations are identical, though, has to be decided once and for all. So the strategy of identifying the P- and Q-events iff both P and Q are efficacious leads to inconsistent results: they *can't* be the same

With the exclusion principle neutralized, the application to mental causation is anticlimactic. As a rule, determinates are tolerant, indeed supportive, of the causal aspirations of their determinables. Why should it be different, if the determinate is physical and the determinable mental? Inferring the causal irrelevance of, say, my *dizziness*, from the causal sufficiency of its physical basis, is not appreciably better than rejecting the redness as irrelevant on the ground that all the causal work is accomplished already by its determinate scarlet. Or, if someone thinks it *is* better, then she owes us an explanation of what the metaphysically important difference is between the cases. That there is a conceptual difference is granted, but it is not to the point; there is no conceptual entailment either from the tea's micromechanical condition to its high temperature, yet this occasions little skepticism about the role of the tea's temperature in its burning my tongue. If there is a metaphysical difference, then someone should say what it is, and why it matters to causation.

5.

According to our guiding principle (**Δ**) for property determination, P determines Q iff to possess the one is to possess the other, not *simpliciter*, but in a certain way. But this way of putting things comes naturally, too, in connection with particulars, and especially events. If p is the bolt's *suddenly* snapping, for example, and q is its snapping *per se*, then for p to occur is for q to occur in a certain way, namely suddenly; and my *slamming* the door consists in my shutting it, not *simpliciter*, but with significant force.[33] This suggests the possibility of a determination relation for events:

(**δ**) p determines q iff: for p to occur (in a possible world) is for q to occur (there), not *simpliciter*, but in a certain way.[34]

---

event, because there are effects (the glass's shattering) to which only P is relevant; at the same time they *must* be, to accommodate effects (the neighbor's turning up her hearing aid) to which Q is relevant too.

[33]Here and throughout 'events' are event tokens, not types; my slamming the door is something that happens at a specific time, in a specific place, and in a particular way.

[34]Where this is understood fairly generally, so that, for example, Poindexter's lying to Congress is his speaking to Congress in a certain way, to wit falsely.

If the relation can be made out, then in addition to the examples mentioned, Icarus's flying too near the sun determines his flying *per se*, Brutus's killing Caesar determines his stabbing Caesar,[35] Gödel's discovering the incompleteness of arithmetic determines his realizing that arithmetic was incomplete, and so on indefinitely.

There is a complication. Determination involves the idea that the requirements associated with one thing include the requirements associated with another; and although properties are requiremental on their face, particulars are not. Hence the need for a notion of individual essence.

By a thing's *essential* properties, I mean those it cannot exist without. And its *essence* is a certain selection of its essential properties. But which essential properties does it make sense to include? The simplest proposal, obviously, would be to include *all* of them. For two related reasons, though, that won't do. Naively, the "what-it-is" of a thing—its identity and kind—should be *in virtue of* its essence. Yet if identity- and kind-properties are allowed into essences, this requirement becomes quickly trivialized: a thing does not get to be identical to Brutus's stabbing Caesar, or of the kind *stabbing*, by having the property of so being, but by having certain *other* properties and by their dividing along appropriate lines between essential and accidental. Second, the essence of a thing is supposed to be a measure of what is *required* in order to be that thing. Thus if more is required to be y than to be x, this should be reflected in an inclusion relation between their essences. The problem is that identity-properties, kind-properties, and the like are liable to disrupt these inclusion relations. Allowing *identity-with-x* into x's essence precludes the possibility of a y whose essence includes everything in x's essence, and more besides; and the effect of allowing x's kind into its essence is to kill the chances for a thing y whose essence exceeds x's by properties which things of that kind possess at best accidentally.[36]

Both problems have the same solution: essences are to be drawn from a pool of properties such that any particular such property's

[35]Killings need not be stabbings, and Brutus could have killed Caesar without stabbing him; but this *particular* killing, I assume, could not have occurred except by way of the associated stabbing (this is important if the killing is to be a determination of the stabbing).

[36]For example, to *stabbings*, unlike *killings*, it is not essential that someone die.

modal status—essential or accidental—is without undue prejudice to the modal status of the others. Dubbing these the *cumulative* properties, x's *essence* will be the set of cumulative properties that it possesses essentially. When q's essence is a subset of p's essence, p is said to subsume q ($p \geqslant q$); and p *determines* q ($p > q$) when the inclusion is strict.[37]

Explaining determination by essence has three points in its favor: it fits the intuitive examples; it supports the analogy with property determination; and it predicts the principle that p determines q only if for p to occur is for q to occur in a certain way. Take the example of Gödel's *discovering*, versus his simply *realizing*, that arithmetic was incomplete. Though identical on some accounts, there is in fact a subtle difference between them. Speaking first of Gödel's *realizing* that arithmetic was incomplete, this *could* have been the realization of a result already widely known (in that case, it would not have made Gödel famous). To Gödel's *discovering* arithmetic's incompleteness, though, some degree of priority is essential. Otherwise one could ask, would it still have made Gödel famous, if incompleteness had been common knowledge? But this is like asking, of Brutus's killing Caesar, what Caesar would have done to Brutus if he had not died of it. So the essence of Gödel's discovering that arithmetic was incomplete *adds* something to the essence of his realizing that it was.

For the analogy with property determination, we need a distinction: a property is *categorical* if its possession by a thing x at a

[37]Here is the basic condition on cumulative properties stated more formally: (κ) for all x, for all possible worlds w, for all sets S of cumulative properties [x exists in w and possesses there every member of $S \leftrightarrow$ there exists in w an $x^+ \geqslant x$ to which every member of S belongs essentially]. To see how this works to exclude identity properties, suppose that x possesses some cumulative P accidentally in some world w where it exists. If *identity-with-x* were cumulative, by (κ) there would be an x^+ in w to which *identity-with-x* and P were both essential—a contradiction, since nothing can be both identical to x and essentially possessed of a property which x possesses only accidentally. Likewise for kind-properties: if x is accidentally P and of such and such a kind, it will normally be impossible to strengthen x into an x^+ still of that kind but possessing P essentially. Thus, no *person* is *essentially* born on a certain day, no *stabbing* is *essentially* fatal, no *landslide* is *essentially* between nine and ten seconds long, and so on. (Terminological note: subsumption is called 'refinement' in "Identity, Essence, and Indiscernibility," *Journal of Philosophy* 84 [1987]: 293–314, and 'strengthening' in "Cause and Essence," *Synthese* [1992].)

possible world is strictly a matter of x's condition in that world, without regard to how it would or could have been; other properties, for example counterfactual and modal properties, are *hypothetical*.[38] This gives the idea of categoricity, but as a definition it would be circular. To see why, suppose it is a categorical property of this piece of wax to be spherical. How can this depend on the wax's condition in other worlds? In a way, though, it does, for the wax cannot be spherical in this world without being possibly spherical in every other world it inhabits. More generally, sensitivity to its possessors' *hypothetical* characteristics in other worlds should not make a property noncategorical, or *no* properties will be categorical. What we *meant* to say, it seems, is that a property is categorical iff it attaches to its objects regardless of how they would or could have been in *categorical* respects. And now the circularity is apparent.

Luckily the categorical properties can be approached from another direction. When p *subsumes* q, their difference (if any) comes down ultimately to the fact that they possess different of their shared properties essentially. Such a difference is *merely* hypothetical if any difference is; so

(γ) C is categorical only if: necessarily, for all p and q such that $p \geq q$, p has C iff q does.

This, although only a necessary condition on categoricity, is all that the announced analogy requires.[39] For it entails that in worlds where both exist, the subsuming particular p and the subsumed q

[38]More familiar are the notions of an *occurrent* property: one whose possession by a thing at a time is insensitive to how matters stand at other times; and an *intrinsic* property: one which a thing possesses wholly in virtue of how it is in itself, irrespective of what goes on around it. Within limits we can think of categoricity as standing to the modal dimension as occurrence stands to time and intrinsicness to space (see "Identity, Essence, and Indiscernibility," and "Intrinsic, Occurrent, Categorical," manuscript).

[39]Assuming that the logical space of particulars is *full* in a sense I discuss elsewhere, the stated condition is sufficient also ("Identity, Essence, and Indiscernibility," secs. 4 and 5). Fullness is a sort of plenitude principle whose point is to ensure that there are particulars enough to witness the hypotheticality of every hypothetical property; that is, that for each

are categorically indiscernible, or as I will say *coincident*. And since p cannot exist *without* q[40] (the bolt's suddenly snapping is impossible without its snapping) we have:

(μ) $p \geqslant q$ only if: necessarily, if p exists, then q exists and is coincident with p.

This divides into two subconditions, according to whether p is identical to q or determines it.[41] By Leibniz's Law, or a double application of (μ),

---

hypothetical H, there exist in some possible world $\geqslant$-related p and q such that H attaches to exactly one of them. To illustrate, part of the assumption is that for any particular q and any non-empty set W of worlds in which it exists, there is a $p \geqslant q$ which exists in the W-worlds exactly. Now suppose we agree that to be, say, flexible, a thing must be at least capable of flexing, that is, it needs to flex in at least some worlds. By fullness, any flexible q, provided only that there are worlds in which it never flexes, will have a determination p which metaphysically *cannot* flex. This shows that flexibility is hypothetical. (Some say that if dispositional properties are hypothetical, then *all* properties are, for it is essential to every property, however categorical it might otherwise seem, to confer on its possessors correlative causal dispositions, for instance, flexibility, corrosiveness, visibility. But the idea that even seemingly categorical properties are essentially disposition-conferring is, in the context of the fullness assumption, quite implausible. For instance, it detracts not at all from a thing's actual-world *roundness* to restrict or otherwise adjust its counterfactual career, but its dispositions can be varied almost at will by the same operation. What *might* be essential to roundness is to confer appropriate dispositions on particulars meeting *further hypothetical conditions*, conditions aimed at ruling out unusual hypothetical coloration such as we saw above. Yet since roundness "entails" these dispositions only over its hypothetically *ordinary* possessors, the objection is analogous to the following: no *ordinary* thing moves discontinuously; so being at such and such a location at a given time "entails" the non-occurrent property of not being at every *other* time a million miles away; so, location properties are not occurrent!)

[40]That is, if $p \geqslant q$, then necessarily if p exists then so does q. *Proof*: Run (κ) from right to left with S = the empty set. (Another proof uses the assumption that x exists in w iff x's essence is satisfied there, that is, something possesses there all its member properties: w contains p only if p's essence is satisfied in w only if q's smaller essence is satisfied in w only if w contains q.)

[41]That these exhaust the possibilities is not trivial; but it can be proven from (μ) and the assumption (σ) that distinct particulars either exist in different worlds or are noncoincident in some world where they exist together. *Proof*: It suffices to show that $p = q$ if they have the same essence.

(ι) $p = q$ only if: necessarily, p exists iff q exists, and if existent, they are coincident.

When p determines q, the condition holds in one direction only:

(δ) $p > q$ only if:
(i) necessarily, if p exists, then q exists and is coincident with p;
(ii) possibly, q exists and p does not exist.[42]

That we get these analogues for particulars of (I) and (Δ) is the second attraction of using essence to explain determination.

Now for the fact that reflects most favorably on the essence approach: that it predicts **(δ)**'s intuitive description of determination. From (δ) we know that a determinate p exists in some, though not all, of the worlds where its determinable q is found. But how does p decide in *which* of these q-worlds to put in its appearances? For instance, what separates the worlds in which the bolt's suddenly snapping accompanies its snapping *per se* from those in which it does not? In the former worlds, presumably, the snapping is sudden; and as it turns out, this answer holds good in general:

(ϵ) $p > q$ only if: necessarily, p exists iff q (both) exists and exemplifies the difference S between its own essence and p's larger essence.[43]

Mirabile dictu, this is just what **(δ)** *says* about determinates and their determinables: for p to occur is for q to occur, not *simpliciter*, but S-ly.

Suppose they do. Then each subsumes the other. By (μ), they exist in the same worlds and are coincident in all of them. By (σ), $p = q$.

[42] *Proof*: (i) is immediate from (μ) and the fact that determination entails subsumption. (ii) If p existed in every world in which q did, then by (μ) and (i) they would exist in the same worlds and be coincident in all of them. Given (σ) that would make them identical, contrary to the assumption that p determines q.

[43] *Proof*: Suppose that q exists in a world w and exemplifies S there. By (κ)'s left-to-right direction, there exists in w a $q^+ \geqslant p$; it follows from (μ) that p exists in w. For the converse, run (κ) from right to left with S as before.

6.

Identicals are indiscernible; so an argument that mental events have different essential properties from physical events is an argument that they are not identical. According to one popular line of thought, this essential difference can be established in the following simple form: only mental events possess mental properties (e.g., phenomenal and content properties) essentially. Thus Kripke:

> Let '*s*' name a particular pain sensation, and let '*b*' name the corresponding brain state, or the brain state some identity theorist wishes to identify with *s*. *Prima facie*, it would seem that it is at least logically possible that *b* should have existed (Jones's brain could have been in exactly that state at the time in question) without Jones feeling any pain at all, and thus without the presence of *s*.[44]

Prima facie, Kripke says, *b* could have occurred without there being any pain, and presumably he would say the same about other physical events *p* and mental properties. Unless these *prima facie* appearances can be overcome, mental properties are at best accidental to physical events.

Are these really the *prima facie* appearances, though? Remember that all it takes for *p* to have a mental characteristic essentially is for its essential physical properties to necessitate one—and that the dominant modal intuition in recent years has been that mental properties *supervene* on physical properties and so are necessitated by them *all the time*.[45] Someone might of course ask why any physi-

---

[44]*Naming and Necessity*, 146, with inessential relettering. Note that if "logically possible" is taken literally, as covering everything permitted by logic, then even identicals can differ in what is logically possible for them—for example, it is logically possible that Hesperus, but not that Phosphorus, should exist in Phosphorus's absence. Obviously this would make logical possibility useless in applications of Leibniz's Law; so I assume that Kripke is using "logical possibility" for metaphysical possibility.

[45]Two remarks. First, the point of calling this an *intuition* is that Kripke's argument might be read as objecting to supervenience itself (*Naming and Necessity*, 155). So read, the argument assumes that the weight of modal intuition favors the antisupervenience position. This I deny; there are many reasons for supervenience's popularity, but one, surely, is its enormous modal intuitiveness. (Such antisupervenience intuitions as may exist

cal *p* should have the mentally consequential *kind* of physical property, but this is easily explained. Consider the bearing of supervenience on *mental* events: for each of *m*'s mental properties, supervenience assigns it a necessitating physical property. But it is hard to think what *m*'s physical properties could be if not those of some physical event *p* which subserved it. Thus, among *p*'s physical properties are some with *m*'s mental properties as necessary consequences. Only if *p* somehow managed to have *all* of these physical properties contingently could it avoid having at least some mental properties essentially.

Instead of insisting that *p* has *no* essential mental properties, perhaps the token dualist should say that it doesn't have *all* the essential mental properties of its alleged mental identical. Here is a bad way to argue for that result: since no mental event is physical, *p* lacks mental kind-properties, for example, being of the kind *after-image*, *sensation*, or indeed *mental*; therefore it doesn't have these properties essentially. Dialectically, of course, this begs the question against the token identity theory. But there is a deeper problem: it says nothing about what *makes* a mental event *m* different from a physical event *p*, to be told that only the former is (essentially) mental, or of some specific mental kind. Mental events are mental rather than physical not because mentality is essential to them alone, but because of some *prior* fact about them—the sort of fact that essences were designed to capture. Thus *m*'s essential mental advantage over *p*, if it exists, should be that its *essence* contains mental properties beyond those in *p*'s essence.

Yet supervenience opposes this weakening of the essential mental advantage view as much as the original. The reason is this. Every

---

I would hope to explain away in the manner of note 24.) Second, someone might complain that "the dominant intuition" is only that the mental characteristics of *objects*, or perhaps *worlds*, are necessitated by their physical properties; *events* are another story. Otherwise supervenience entails, as it surely should not, that every mental event is a physical event. But the objection assumes that events with mental properties are thereby mental events; and I am in the process of questioning whether even *essential* mental properties are enough to make an event mental. (For the idea that strong supervenience presupposes token identity, see Haugeland, "Weak Supervenience," *American Philosophical Quarterly* 19 [1982]: 93–103, and Kim, "Supervenience for Multiple Domains," *Philosophical Topics* 16 [1988]: 129–50.)

mental property M_k in m's essence is backed by a necessitating physical property P_k; and as before, these physical properties attach also to some realizing physical event (this time called q). Even if some or all of the P_ks are only accidental to q, we can imagine a more determinate physical event p to which they are all essential. But then p has essential physical properties to necessitate every mental property in m's essence; and it follows that these mental properties are in p's essence too. Not only does this rule out an essential mental advantage for mental events, it puts us in sight of an intriguing parallel between the ways that mental events and properties relate to their physical underpinnings. For assuming that p can be chosen determinate enough to essentially possess such few *non*mental properties as might be found in m's essence, we have

(s) Whenever a mental event m occurs, there occurs also a subsuming physical event p, that is, a physical event whose essence includes m's essence[46]

—an analogue for events of the supervenience thesis.

From (s) it is clear that if there is an essential difference between mental events and physical ones, it is *not* that physical events' essences are mentally impoverished. Instead, I suggest, it is the other way around: the essences of *mental* events are *physically* impoverished. For those who believe, with Descartes, that their mental lives could have proceeded just the same in a wholly immaterial world, this hardly requires argument.[47] Events which can occur in such a world presumably have *none* of their physical properties essentially. But Cartesian dualism is only the most dramatic expression of a thought which seems probable in any case, namely, that in com-

[46]Notice what (s) doesn't say: that every property *essential* to m is essential to p. For all we know so far, no mental event is physical; in that case m's mental identity- and kind-properties are not properties of p at all.

[47]Of course, whoever accepts supervenience in form (S) will find the Cartesian hypothesis hard to swallow, for (S) implies that in, and across, immaterial worlds, everyone is thinking exactly the same thing! This has led some authors (e.g., David Lewis) to seek more permissive interpretations of supervenience.

parison with their physical bases, mental phenomena are exceedingly modally elastic.[48]

Take for example the pain sensation *s*, and the underlying brain event *b* whose identity with *s* is in question; and grant the identity theorist that *b* at least subsumes *s* and so necessitates it. The problem is that as *b* takes on the degree of essential physical detail that this requires, it becomes intuitively irresistible that the pain is possible even in *b*'s absence. Something like this is Kripke's second argument against the identity theory:

> [*B*]*eing a brain state* is evidently an essential property of *b* (the brain state). Indeed, even more is true: not only being a brain state, but even being a brain state of a specific type is essential to *b*. The configuration of brain cells whose presence at a given time constitutes the presence of *b* at that time is essential to *b*, and in its absence *b* would not have existed. Thus someone who wishes to claim that the brain state and the pain are identical must argue that the pain could not have existed without a quite specific type of configuration of molecules.[49]

Prima facie, it seems obvious that the pain could still have occurred, even if that specific arrangement of molecules hadn't, and as Kripke says, the *prima facie* appearances aren't easily defeated.[50]

---

[48]This is a particular theme of Richard Boyd, "Materialism Without Reductionism," in *Readings in the Philosophy of Psychology*, vol. 1, ed. N. Block (Cambridge: Harvard University Press, 1980), 67–106.

[49]*Naming and Necessity*, with inessential relettering.

[50]"Granted that *a* pain could still have occurred in the absence of that molecular configuration, what makes you think that it is the same pain that occurred actually?" Among the lessons of *Naming and Necessity* is that to find a thing *x* capable of existing in some counterfactual condition, one imagines this *directly*—as opposed to imagining something *y* in that condition whose transworld identity with *x* must then be established. This is crucial if imaginability is to be a source of knowledge about *de re* possibility. For (i) having imagined *y* in the indicated condition, verifying that *y* is *x* requires appeal to transworld identity criteria which, if they are available at all, are typically *more* controversial than the *de re* attributions they are called on to support; and (ii) without reliance on direct *de re* imagination there would be no way to justify these criteria in the first place. Stripped then of its reference to transworld identity, the question is, Is *m* really imaginable in the absence of *b*, or is the only imaginable scenario one in which a distinct if similar pain occurs in *b*'s absence? Here I can do no more than echo Kripke in claiming the former intuition. Such intuitions are of

But if the molecular arrangement is essential to b alone, then b's essence is physically richer than s's essence. Therefore b subsumes s *properly*; and this, extended across mental events in general, gives an analogue for particulars of the multiple realizability thesis:

(m) For every mental event m, and every physical event p which subsumes m, p subsumes m properly and so determines it.

Token dualism follows: if m were identical to p, then p would subsume m; hence by (m) it would determine m, contrary to their assumed identity.

Drawing these various threads together, we find that the relation between mental and physical events effectively duplicates that of mental to physical properties. Whenever a mental event m occurs, (s) guarantees a subsuming physical event p, which by (m) is not identical to m but determines it. Thus with every mental m comes a determining physical p.[51] Since for p to occur is just for m to occur in a certain physical way, the converse is trivial; so we can say that

---

course defeasible by reference to unnoticed complications, but they are *prima facie* credible and the burden of proof is on the critic (see Boyd, "Materialism Without Reductionism," for pertinent thought experiments, and note 24 for the defeasibility of modal intuition). On a deeper level, perhaps the objection reflects not any particular attachment to a picture of mental events as bound to their physical underpinnings, but a more general malaise attending *all* modal thinking about events. Whereas objects fall into more or less settled kinds, which then guide us in our assessment of what counterfactual changes they will tolerate, with events our commonsense sortal apparatus is relatively primitive and modally inarticulate; that something is a pain, or an explosion, tells us enormously less about its possibilities than that it is a person or a ship. Hence our admitted squeamishness about events' potential for contrary-to-fact behavior—which hardens all too easily into the positive thesis that that potential is extremely limited (i.e., that events are inherently modally inflexible). This last, though, is surely an overreaction. What the squeamishness really signifies is the inadequacy of everyday event-sortals to the task of identifying just which of various coincident-but-hypothetically-different items one has in mind. Small wonder, then, if the identificatory task falls partly to the *de re* modal attributions themselves; and some of the more dogmatic-sounding attributions in the text may seem less so when understood in this spirit: as partial specifications of their subject matter rather than as attempts to describe an already singled-out particular.

[51]This may seem doubtful, if one insists on seeing p as (i) a localized brain event, (ii) capable of occurring in isolation from anything like its

(d) A mental event m occurs iff some physical determination p of m occurs.

This is our analogue for events of the mental/physical determination thesis for properties.

7.

Haven't we now made mental events causally irrelevant? By the exclusion principle, m can influence an outcome only to the extent that p leaves that outcome causally undecided. Results which p causally guarantees, therefore, it renders insusceptible to causal influence from any other source, m included. Assuming, for example, that all it took for me to wince, clutch my brow, and so on, was my antecedent physical condition, everything else was strictly by

---

actual neural context. Imagine a C-fiber stimulation, b, and a pain sensation, s, with the following properties. First, they are both occurring in me right now; second, b *could* have occurred in isolated C-fibers afloat in agar jelly; third, had b occurred in the latter environment, s would not have accompanied it. Then since determination entails necessitation, b does not determine s. The moral is that (i) and (ii) ask too much. Most mental events m seem not to be localizable in any specific portion of the brain; determination entailing coincidence, their physical determinations p will not be localizable either (thus p might be the event of my falling into a certain overall neural condition). Perhaps no mental event is localizable, but if m is an exception, its physical determination p will have a partly extrinsic essence (thus p might be my C-fibers' firing in normal neural surroundings). So-called "wide content" mental events raise related but different problems which I don't discuss. Possibly they will have to be allowed as exceptions to the physical/mental determination thesis; in that case, the paper should be read as defending the causal potency of *other*-than-"wide content" mental events. Two remarks, though, to put this in perspective: First, it is controversial how often such events are genuinely efficacious, in particular because their "narrow" counterparts seem ordinarily to be more commensurate, in the sense of section 8, with their supposed effects (see J. Fodor, *Psychosemantics* [Cambridge: MIT Press, 1987], chap. 2 and "A Modal Argument for Narrow Content," *Journal of Philosophy* 88 [1991]: 5–26). Second, determination is only the most obvious of a number of intimate identity-like relations equally unsupportive of the "x_1 was sufficient, so x_2 was irrelevant" reflex. Neither of Beamon's outjumping the competition and his jumping 29′ 2¼″ determines the other; but nobody would think the latter irrelevant to his being awarded the gold medal because the former was sufficient (see J. Heil and A. Mele, "Mental Causes," *American Philosophical Quarterly* 28 [1991]: 49–59).

the way. Since my headache is a different thing from its determining physical basis, it is not a *bona fide* causal factor in my headache behavior.

By now the deficiencies of this line of argument must be apparent. Suppose that we think of the exclusion principle as saying that for every irreflexive relation R (every "form of nonidentity"), and every R-related pair x and x^*, x's causal sufficiency for an effect entails x^*'s causal irrelevance. Though there may be irreflexive relations R whose relata *do* contend for causal influence as the principle says, for many Rs this competition arises only sometimes, and for others it *never* arises. Ironically, R = causation is a case in point. Let x be causally sufficient for y. Then taken at its word, the exclusion principle predicts that y owes nothing to the causal intermediaries by which x brings y about. When R is causation's converse, the prediction is different but still absurd: events causally antecedent to x can claim no role in y's production.[52] Of course, the case that interests us is R = the determination relation. Remember Archimedes' excited outburst on discovering the principle of displacement in his bath. Assuming that his shouting "Eureka!!" was causally sufficient for his cat's startled flight, nobody would think that this disqualified his (simply) shouting from being causally relevant as well. And it would be incredible to treat Socrates' *drinking* the poison as irrelevant to his death, on the ground that his *guzzling* it was causally sufficient.

Thinking of causal influence as something that an effect's would-be causal antecedents compete over in a zero-sum game, the exclusion principle looks not unreasonable. If the causally sufficient antecedent monopolizes *all* the influence, then the others are left with none. To judge by the examples, though, causation is not like that: rather than competing for causal honors, determinables and their determinates seem likelier to share in one another's success. Again the application to mental and physical events is anticlimactic. Unless an arbitrary exception is to be made of them, it is no argument at all for the causal irrelevance of, say, a sensation that its

[52]Goldman, "The Compatibility of Mechanism and Purpose," and Kim, "Mechanism, Purpose, and Explanatory Exclusion" make related observations.

occurring in some specific physical way was causally sufficient.[53] With events as with properties, physical determinates cannot defeat the causal pretensions of their mental determinables.[54]

8.

To this point our position is wholly negative: for all that the exclusion argument shows, mental phenomena *can* be causally relevant compatibly with the causal sufficiency of their physical bases. It is a further question whether they *will* be in any particular case. And even if some mental antecedent *is* causally relevant, it is a further question yet whether it actually *causes* the effect.

Notice some important differences between causal relevance and sufficiency, on the one hand, and causation, on the other: x can be causally sufficient for y even though it incorporates enormous amounts of causally extraneous detail, and it can be causally relevant to y even though it omits factors critical to y's occurrence. What distinguishes causation from these other relations is that causes are

[53]Lately there has been a tendency to argue that p's causal sufficiency for an effect, though it does not *directly* entail m's irrelevance, limits m's role to that of a causal overdeterminant at best (see note 6); that m is indeed irrelevant then emerges from the fact that the effect is not overdetermined. With as much or little plausibility, one could argue that Ella's singing at over 70 db was irrelevant to the glass's breaking, since the latter was causally guaranteed, but not overdetermined, by her singing at 80 db exactly.

[54]Suppose that causal sufficiency is read in some fairly demanding way, say, as requiring the strict nomological impossibility of x's occurring without y's doing so. Then no physical event p with hopes of determining a mental event m is likely to be itself causally sufficient for m's apparent effect y. For p can determine m only if they are the same size, and nothing that small—assuming anyway that its essence is not unconscionably extrinsic—can nomologically guarantee any but the most trifling and immediate results. Let it be granted, then, that p is not causally sufficient for y; that honor falls instead to a spatially more extensive physical event p', whose occurrence essentially requires, in addition to p's occurring, that the surrounding physical conditions be approximately as they are in fact. This affects the question of m's causal potency, *only* if there is more causal rivalry between m and p' than we found between m and p (namely, none). But, how could there be? What dispelled the illusion of rivalry between m and p was that p's occurrence consisted, in part, in m's occurrence, and that is as true of m and p' as it was of m and p: for p' to occur is for m to occur in a certain physical way, and in a certain physical environment. So p' poses no greater threat than p to m's causal aspirations.

expected to be *commensurate* with their effects: roughly, they should incorporate a good deal of causally important material but not too much that is causally unimportant. And this makes causation special in another way. Although determinables and determinates do not compete for causal *influence*, broadly conceived as encompassing everything from causal relevance to causal sufficiency, they *do* compete for the role of *cause*, with the more commensurate candidate prevailing. Now I argue that the effect's mental antecedents often fare *better* in this competition than their more determinate physical bases.[55]

Inspiring the commensuration constraint is a certain platitude: the cause was the thing that "made the difference" between the effect's occurring and its not. Had the cause been absent, the platitude seems to say, then (i) the effect would have been absent too, but (ii) it *would* have occurred if the cause had. Thus effects are *contingent* on their causes:

(C) If x had not occurred, then y would not have occurred either;[56]

and causes are *adequate* for their effects:

(A) If x had not occurred, then *if it had*, y would have occurred as well.[57]

---

[55]To keep things simple, I'll focus on mental events; there is a related story about mental properties.

[56]For definiteness, we interpret would-counterfactuals Stalnaker's way: 'if it had been that P, then it would have been that Q' is true iff Q is true in the P-world best resembling actuality; where it is indeterminate which P-world that is, the condition must hold on all admissible ways of resolving the indeterminacy. Might-counterfactuals, 'if it had been that P, then it might have been that Q', are true just in case their associated would-counterfactuals, 'if it had been that P, it would have been that not-Q', are *not* true. Equivalently, a might-counterfactual holds iff on at least one admissible selection of closest P-world, the closest P-world is a Q-world. (See Lewis, "Counterfactuals and Comparative Possibility," and Stalnaker, "A Theory of Conditionals" and "A Defense of Conditional Excluded Middle," all in *Ifs*, ed. W. L. Harper, R. Stalnaker, and G. Pearce [Dordrecht, The Netherlands: D. Reidel, 1981].)

[57]Rasmussen, "Ruben on Lewis and Causal Sufficiency" (*Analysis* 42

Without mentioning determination explicitly, these conditions do nevertheless discover causal differences between unequally determinate events. Suppose we stipulate that it contributed nothing to Socrates' demise that he guzzled the hemlock rather than simply drinking it. Then Xanthippe is mistaken when, disgusted at Socrates' sloppy habits, she complains that his *guzzling* the hemlock caused his death. Assuming that the drinking would still have occurred, if the guzzling hadn't, (C) explains the error nicely. Even without the guzzling, the death would still have followed on the drinking. So while Socrates' death may have been contingent on his drinking the hemlock, it was *not* contingent on his guzzling it.[58]

Here the contingency condition exposes an overly determinate pretender; sometimes, though, the pretender's problem is that it is not determinate enough. Safety valves are designed to open quickly under extreme pressure, thus easing the burden on the equipment upstream. This particular valve has begun to operate as advertised when a freak molecular misalignment stiffens the mechanism; this decelerates the opening to just past the point of endurance and the boiler explodes. Assuming that the explosion does *not* result from the valve's opening *per se*, I ask why not. Because the contingency condition is violated? But we can arrange it so that the explosion *was* contingent on the opening, say, by stipulating that if the opening had not occurred, rather than the boiler's exploding the connecting pipe would have burst. Adequacy does better: given the unlikelihood of the molecular mishap, had the

---

[1982]: 207–11) contains the only explicit reference to (A) that I have seen. There it is argued, fallaciously I think, that (A) follows from (C) on the assumption that x and y actually occur. Another erroneous criticism, encountered mostly in conversation, is that (A) is trivial given just the occurrence of x and y: (A) is true iff y occurs in the nearest x-containing world w to the nearest x-omitting world v to actuality; but since x actually occurs, the nearest x-containing world w to v is the actual world, which contains y by hypothesis. This forgets that w is the actual world only if no x-containing worlds are nearer to v than the actual world is, and that some are bound to be if, as seems likely, the actual world sits in the interior of a neighborhood of x-containing worlds. (What *does* follow trivially from the occurrence of x and y is the condition that if x had occurred, so would have y; this is why we use (A) despite its greater complexity.)

[58]David Lewis puts contingency to similar use in his "Events," in *Philosophical Papers*, vol. 2 (Oxford: Oxford University Press, 1986).

opening failed to occur, it might easily have been quicker if it had.[59] Speaking then of how things *would* have been if not for the opening, it cannot be said that, *were* it to have occurred, it would still have brought the explosion in its wake.

Important as they are, contingency and adequacy capture the commensuration intuition only partly. Imagine that Socrates, always a sloppy eater, had difficulty drinking without guzzling, to such a degree that if the guzzling hadn't occurred, the drinking wouldn't have either. Then Socrates' death *was* contingent on his guzzling the hemlock; and so more than contingency is needed to explain why it was not the effect of his doing so. Intuitively, it appears that not *all* of the guzzling was needed, because there occurred also a lesser event, the drinking, which would still have done the job even in the guzzling's absence. By hypothesis, of course, without the guzzling this lesser event would not have taken place; but that doesn't stop us from asking what would have happened if it had, and evaluating the guzzling on that basis. Suppose we call x *required* for y just in case

(R) For all $x^- < x$, if x^- had occurred without x, then y would not have occurred.

Then what disqualifies the guzzling is that, given the drinking, the death did not require it.

Symmetry considerations suggest the possibility of a condition complementary to (R), and a variation on the valve example shows that one is in fact needed. Imagine that the mechanism stiffens, not extemporaneously as above, but because of a preexisting structural defect that would have decelerated the opening in any case. Presumably this means that if the opening had not occurred, it would

[59] I emphasize that the decelerating stiffness sets in only *after* the opening gets under way because I want it to be clear that *that very opening* could have been less protracted (as opposed to: a slower opening could have occurred in its place). To deny this would be to hold that the opening, once begun, *could not* have continued apace, that is, that the approaching deceleration was essential to it. As for the further claim that it *might* have been less protracted, suppose if you like that indeterminism holds, and that the misalignment's objective probability, conditional on preceding events, was extremely low. (The relation between 'would' and 'might' is described in note 56.)

still have been protracted if it had, and the explosion would still have ensued. Since now the opening *is* adequate for the effect, the problem with taking it for the cause lies elsewhere; and the obvious thought is that the effect required something more. Thus define x as *enough* for y iff no more than x was required:

(E) For all $x^+ > x$, x^+ was not required for y.

Because the valve's *slowly* opening was required for the explosion, its opening *per se* was not enough; and that is why it was not the cause.

When all of the conditions are met—that is, y is contingent on x, and requires it, and x is adequate, and enough, for y—x will be called *proportional* to y. Without claiming that proportionality is strictly necessary for causation,[60] it seems clear that faced with a choice between two candidate causes, normally the more proportional candidate is to be preferred. Which of the contenders proportionality favors depends, of course, on the effect in view; Socrates' drinking the hemlock is better positioned than his guzzling it to cause his death, but relative to other effects proportionality may back the guzzling over the drinking.

More to the present point is the following example: I arrive on your doorstep and, rather than knocking, decide to press the buzzer. Epiphenomenalist neuroscientists are monitoring my brain activity from a remote location, and an event e in their neurometer indicates my neural condition to be such and such. Now, like any mental event, my decision m has a physical determination p, and the question arises to which of these the neurometer reading e is due. The scientists reason as follows: Because the neurometer is keyed to the precise condition of his brain, e would not have occurred if the decision had been taken in a different neural way, in particular if it had occurred in p's absence. So m was not enough for

[60]Because of the problems of preemption, overdetermination, and so on, strictly necessary conditions on causation are extremely hard to find. As far as I know, philosophers have not succeeded in turning up even a single one, beyond the trivialities that cause and effect should both occur and be suitably distinct (see Lewis, "Causation," with "Postscripts," in *Philosophical Papers*, vol. 2).

e;[61] *p* on the other hand looks *roughly* proportional to *e* and so has the better claim to cause it. Another triumph for epiphenomenalism!

Everything is all right except for the last step. What is true is that *this* mental event did not cause *that* effect. But who would have thought otherwise? When an effect depends not simply on an event's occurring, but on its occurring in some specific manner, one rightly hesitates to attribute causation. Taking the meter reading to result from my decision would be like attributing Zsa Zsa's speeding citation to her driving through the police radar *per se*, or the officer's abrasions to her touching his face.

Then when *do* we attribute effects to mental causes? Only when we believe, I can only suppose rightly, that the effect is relatively insensitive to the finer details of *m*'s physical implementation. Having decided to push the button, I do so, and the doorbell rings. Most people would say, and I agree, that my decision had the ringing as one of its effects. Of course, the decision had a physical determination *p*; but, most people would also say, and I agree again, that it would still have been succeeded by the ringing, if it had occurred in a different physical way, that is, if its physical determination had been not *p* but some other physical event. And this is just to say that *p* was not *required* for the effect.

Remember that this makes no prediction about what would have happened if the decision had occurred in *whatever* physical way, but speaks only of what transpires in the *nearest* world where its physical implementation was not as actually—the world in which it undergoes only the minimum physical distortion required to put its actual implementation out of existence. Maybe, of course, we were wrong to think that the ringing would still have occurred in that world; if so, then let us hurry to withdraw the assertion that the decision caused it (the real cause is some physically more determinate event). But if not, then our conclusions should be these (where *r* = the doorbell's ringing):

(i) *m* is a counterexample to *r*'s requiring *p* (for *r* would still have occurred, if *m* had occurred without *p*);

[61]Strictly speaking this assumes that *p* was required for *e*—in other words, that *each* of *p*'s determinables, not just *m*, is such that if it had occurred in *p*'s absence, *e* would not have ensued. (For the interpretation of (R) and (E)'s event-quantifiers see "Cause and Essence," sec. 11.)

(ii) p is not proportional to r (since r does not require it);
(iii) p does not cause r (since it is not proportional to r);
(iv) p is not a counterexample to m's enoughness for r (it could be a counterexample only if r required it);
(v) p is not a counterexample to m's proportionality with r (by inspection of the remaining conditions);
(vi) p poses no evident threat to the hypothesis that m caused r.

Here are the beginnings, at least, of a story wherein a mental event emerges as better qualified than its physical basis for the role of cause. I believe that this *kind* of story is enacted virtually wherever common sense finds mental causation.

9.

Indeterministic scruples aside, everything that happens is in strict causal consequence of its physical antecedents. But causally necessitating is a different thing from causing, and the physical has no monopoly on causation. Among causation's prerequisites is that the cause should be, as far as possible, commensurate with its effect; and part of commensuration is that nothing causes an effect which is essentially overladen with materials to which the effect is in no way beholden. This, though, is a condition of which would-be physical causes often fall afoul, thus opening up the market to less determinate events with essences better attuned to the effect's causal requirements. Sometimes, these events are mental; and that is how mental causation happens.

In a "Concluding Unscientific Postscript" to "The Conceivability of Mechanism," Malcolm remarks that

> it is true for me (and for others, too) that a sequence of sounds tends to lose the aspect of speech (language) when we conceive of those sounds as being caused neurologically. . . . Likewise, a sequence of movements loses the aspect of action . . . ;

and he asks, "Is this tendency due to a false picture or misleading analogy?"[62] Many philosophers, anxious to defend the possibility

[62]Malcolm, "The Conceivability of Mechanism," 149.