



## **A Mind Is a Terrible Thing to Waste**

Clark Glymour

*Philosophy of Science*, Vol. 66, No. 3 (Sep., 1999), 455-471.

Stable URL:

<http://links.jstor.org/sici?sici=0031-8248%28199909%2966%3A3%3C455%3AAMIATT%3E2.0.CO%3B2-2>

*Philosophy of Science* is currently published by The University of Chicago Press.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ucpress.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# A Mind is a Terrible Thing to Waste— Critical Notice: Jaegwon Kim, *Mind in a Physical World* \*

Clark Glymour†‡

Departments of Philosophy, Carnegie Mellon University, University of California-San Diego

Jaegwon Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press (1998), 146 pp.

---

Jaegwon Kim's *Mind in a Physical World* is an argument about mental causation that provides both a metaphysical theory and a lucid commentary on contemporary philosophical views. While I strongly recommend Kim's book to anyone interested in the subject, my endorsement is not unconditional, because I cannot make the same recommendation of the subject itself. Considering arguments of Davidson, Putnam, Burge, Block, and Kim himself, I conclude that the subject turns on a variety of implausible but received arguments, and that a useful study of mental causation cannot be divorced from scientific details of cognitive psychology, physics, and neuroscience.

---

**1. Introduction.** One of the many things I like about Jaegwon Kim's *Mind in A Physical World* is that it twice cites Samuel Alexander, the very late author of *Space, Time and Deity*. I do not believe I have read another work that takes the opportunity. The citations made me smile, twice, because when as a freshman at the University of Montana I mentioned the book to my teacher, the late Cynthia Schuster (I was working through the small University library, alphabetically by author;

\*Received April 1999.

†Send requests for reprints to the author, Department of Philosophy, University of California-San Diego, La Jolla, CA 92093.

‡I thank Daniel Wegner of the Department of Psychology, University of Virginia, for helpful discussions and references, and Patricia Churchland, Philip Kitcher, Susan Sterrett, and especially Patricia Kitcher for useful suggestions. I shamelessly borrowed a family joke about the joints from Alison Gopnik. Of course, none of them are responsible for what I have said or how I have said it.

Philosophy of Science, 66 (September 1999) pp. 455–471. 0031-8248/99/6602-0008\$2.00  
Copyright 1999 by the Philosophy of Science Association. All rights reserved.

I think I got as far as Kraft-Ebbing), she told me a story about Alexander. It goes like this:<sup>1</sup>

Having arrived at some academic reception on his bicycle with his ear trumpet—he was quite deaf—Alexander was introduced to an American guest:

“Professor Alexander, this is Dr. Homer White, from the University of Chicago, in America, who is professor of Business Ethics.”

“What?”

“THIS IS DR. HOMER WHITE!”

“Yes, have that, White.”

“FROM THE UNIVERSITY OF CHICAGO!”

“Indeed.”

“IN AMERICA!”

“Right, America.”

“PROFESSOR OF BUSINESS ETHICS!”

“Couldn’t hear you.”

“PROFESSOR OF BUSINESS ETHICS!”

“Sorry, can’t make any sense of it. Sounds just as if you’re saying ‘*business ethics*’.”

I will tell you more about Kim’s new book, eventually. But first a word about the company it keeps.

**2. Company.** If you want to find out what a philosopher is really up to, read an introductory textbook. There, for once, the philosopher faces readers with Samuel Alexander’s temperament, readers whom the philosopher cannot (or should not) presume agree that the enterprise is serious. Kim’s own textbook, *Philosophy of Mind*, is unrevealing; Kim just assumes the reader is with him and plunges in with no apologies, but John Heil’s textbook on the same subject (with the same title and date of issue—sounds just like *philosophy of mind*) may serve as a model. Heil begins with a nice twist on an old saw: A tree falls in the woods. If there is someone to hear it, there is a sound of falling; otherwise there are merely sound waves. So, if someone does hear the tree fall, where is the difference, the experience of the sound? Open up the someone’s brain, examine it all you want, you won’t find the sound, or the experience of the sound. What is the relation between the body and brain of the person and the experience of the sound?

This *seems* like a motivation for intimacy between philosophical and scientific questions: can we measure the intensity of the experience of

1. Susan Sterrett pointed out to me that Daniel Dennett has recently published the same anecdote, which would be bootstrap confirmation of its veracity save for the fact that I’m sure I told it to him some years ago.

sound (we can and have, from early psychophysics on); what are the neural processes that produce the experience of sound (we know some of them); what are the developmental and neural processes that enable us to recognize sounds and distinguish among them and understand them (the questions that began neuropsychology); are the neural structures that sustain those capacities invariant from person to person, and from species to species, and if not, how do they vary (doubtless many people know a lot about this, but not me); can the fact that someone is (or is not) experiencing a sound be predicted from neurological measurements (in some cases it can already); can neurological observations suffice to predict *which* sounds someone is experiencing (looks as if they can); could a transformer be devised that would generate the sound a person is experiencing from measurements of his brain as he is experiencing the sound (we could then observe someone else's experience of the sound, but we are not anywhere near a brain phonograph yet); can artificial systems do the same things, or some of the same things (yes), in the same way (who knows)? And that's just the beginning, off the top of my head.

Not according to Heil. Wrong questions. Not philosophy, and too easy besides:

Many philosophers . . . promote the idea that the philosophy of mind is, or ought to be, one component of what has come to be called cognitive science. Cognitive science includes elements of psychology, neuroscience computer science, linguistics, and anthropology. What has a philosopher to offer the scientists who work in these areas? That is a good question.

Perhaps philosophers can provide some kind of unifying influence, a general picture that accommodates finer-grained assessments issuing from the scientific contributors to cognitive science. This, it would seem, is simply to engage in a kind of attenuated metaphysics. The metaphysics is attenuated to the extent that it excludes traditional ontological concerns, and excludes as well considerations of the bearing of sciences like physics on the ontology of mind.

If I sound skeptical about attempts to assimilate the philosophy of mind to cognitive science, I am. This book is premised on the conviction that the philosophy of mind is continuous with metaphysics as traditionally conceived. The difficult questions that arise in the philosophy of mind—and some would say the difficult questions *tout court*—are at bottom metaphysical questions (1998, 7).<sup>2</sup>

2. In case these arguments fail to convince the skeptical student, Heil has another: the student who drops the subject will be "diminished." So there.

Hilbert's Problems, Fermat's Theorem, the Poincaré Conjecture—not metaphysical, weren't difficult enough for philosophers, had to leave those questions to those lightweights, the mathematicians. Is the density of matter enough to close the space of the universe? How does the brain work? All questions insufficiently difficult to merit the attention of really first-class minds. Nothing for philosophers to do in cognitive science but "attenuated metaphysics": might philosophers propose experiments in cognitive psychology, rectify some of its chaotic methodology, derive unforeseen implications, characterize the value and limits of various kinds of data used in cognitive studies, clarify opaque formulations, ferret out contradictions, correct or provide alternatives to normative models? All things a few professional philosophers have done, but of course not the sort of thing that is *really* philosophy.

Heil's parochialism says what a great deal of mainstream-metaphysical-philosophy-of-mind (hereafter, MMPM) supposes. Philosophical projects are to be walled off from any real use of (or for) mathematics or the sciences. Aside from a bit of formal logic, to be informally used, and the philosophical tradition itself, the philosopher faces the dragons in the labyrinth of metaphysics armed only with words and a vivid imagination. Kim appears to be similarly armed.

Even so, especially in philosophy of mind, various philosophical projects are motivated by developments in what might be called The Scientific Viewpoint, which is pretty much just the idea that everything in and about the world depends completely on physics: same physical arrangement of the world, same everything else.

**3. The Mental Causation Challenge.** On brief reflection, the Scientific Viewpoint may seem inconsistent with our ordinary ways of talking about ourselves and others, and our ordinary ways of justifying or condemning actions. In the Democracy of Ideas, philosophers vote Republican: they aim to save conventional ways of talking and acting in the face of The Scientific Viewpoint. So in recent decades most of the philosophical literature on freedom of the will has aimed to argue that it is O.K. for us to talk and practice more or less as of old. And so, too, with mental causation.

The problem of mental causation is this: We believe, and we want it to be true, that our thoughts, beliefs, desires, plans, hopes, and dreads cause our actions, or at least cause some of what we do. We sophisticates have absorbed Freud; we recognize that not all thought is conscious thought, and we may not want so very much that our unconscious thoughts cause our actions—perhaps we would prefer they not—but we very much want to have reason to believe that our con-

scious thoughts, our deliberations, influence what we do. Otherwise, it seems, what each of us cares profoundly about—his and her own consciousness, and the consciousness of others—is just going along for the ride, mounted ineffectually aboard a zombie, shouting “that-a-way” a bit after the zombie has headed that-a-way, deluded that the zombie takes notice and obeys. Our belief in our own mental powers seems inconsistent with our up-to-date scientific belief that same physics implies same everything else. Because *that* seems to imply that all the real causes of anything—all the conditions and alterations that bring about or sustain other conditions and changes—are physical causes. So the mental doesn’t cause anything, and so, too, the conscious mental doesn’t cause anything. It just goes along for the ride.

The Mental Causation challenge is to show how, nonetheless, it is reasonable to believe in mental causation. There seem to be four strategies, each, of course, with complex variants. Of course.

1. Deny the Premise Strategy: The claim that the mental supervenes on the physical is false.
2. The Humpty Dumpty Strategy: The argument against mental causation is sound (modulo idiosyncratic reconstructions) but for a long time we have been talking very satisfactorily about thoughts as causes, and we plan to continue, and the science really has nothing to do with it.
3. The Functionalist Strategy: Mental states are causal dispositions, implemented or realized by physical states.
4. The Identity Strategy: Some mental entities are identical to physical entities, the very same thing, like Cicero and Tully, or the Morning Star and the Evening Star. If those physical entities can be causes, then so can those mental entities.

I hold that a particular version (*my* particular version, of course) of The Humpty Dumpty Strategy is correct. But the Identity and Functionalist Strategies are where the action now resides in the Mental Causation challenge. For a path through the action, for a clear line of argument that you can read in a long evening, without bombast or rudeness or phony verbiage (or my sarcasm), for that precious intellectual oxymoron—a slim volume of philosophy that flows like a pretty good detective story—read Kim’s *Mind in a Physical World*.

**4. Axiomatics.** *Mind in a Physical World* is set in a problematic framed by two philosophers, Hilary Putnam and Donald Davidson. Putnam’s contribution is two theorems, consequences of assumptions made almost everywhere in MMPM (mainstream-metaphysical-philosophy-

of-mind, remember?) but seldom acknowledged. One rendition of them, mine for the moment, goes like this:

*Cartesian Axiom:* Whatever is imagined is possible.

*Pseudo-Fregean Axiom:* Necessarily, two thoughts have the same content only if they have the same truth value.

*Cartesian Theorem:* If two descriptions of properties can be imagined not to be coextensive, they are not descriptions of the same property.

*Putnam's Theorem:* Mental properties are Multiply Realizable—that is, systems of different chemical and physical composition can exhibit the same arrays of mental properties.

*Putnam-Burge Theorem:* Some mental properties of a person are not identical with any physical properties of the same person.

Putnam's Theorem, which follows straightforwardly from the Axioms supplemented with a little imagining, is the basis of contemporary functionalism.

The proof of the Putnam-Burge Theorem merits some comment. Putnam says he can imagine a world, Twin Earth, in which water is not  $H_2O$ , but something else that plays all (or enough of), of the causal roles of water on Earth. On Twin Earth, Twin Putnam is a duplicate of Putnam to the atom. By the Cartesian Axiom, Twin Earth is possible. On Earth, Putnam is thinking of water, which is  $H_2O$ , and so Putnam has the property that he is thinking of something that is  $H_2O$ . On Twin Earth, Putnam's doppelgänger has the same physical properties as Putnam, and is thinking of water as well, but not of  $H_2O$ , so Twin Putnam does not have the property that he is thinking of something that is  $H_2O$ . So Putnam and Twin Putnam do not have the same mental state. Since it can be imagined that Putnam and Twin Putnam have the same physical properties but not the same mental properties, some mental properties of Putnam are not the same as any of his physical properties. Ditto for anyone else.

Now this proof is obviously troubled, since if water is identically  $H_2O$ , Twin Earth is unimaginable, and even if it isn't, since Putnam's brain is made of  $H_2O$ , Twin Putnam's isn't, and so Putnam and Twin Putnam can't consistently share all of their physical properties. Burge fixed it all up by imagining a person in two different worlds in the very same physical state but in different linguistic communities. That imagining might be made into a proof this way: In one world when she has a pain in the thigh and says "I have arthritis," she is wrong, because arthritis is an inflammation in the joints. In the other, when she has a pain in the thigh and says "I have arthritis," she is right, because in the linguistic community in which she lives in that world, "arthritis" is

used to include inflammations in the muscles. So in one world she has the (mental and semantical) property of having a false thought and in the other a true thought. So by the Pseudo Fregean Axiom the two thoughts are not the same, even though her physical state in which she has them is the same for each thought. QED.

(Perhaps the moral is that if you don't want to get stiff in the joints, stay out of the joints.) This is not, I hasten to add, Burge's literal argument, which depends on specific views about which imaginable changes in the world alter meanings and which alter facts without altering meanings. (The argument above would, by parallel reasoning, yield the conclusion that you are in different mental states when you think truly that you left your keys in the car and when you think falsely that you left your keys in the car.<sup>3</sup>) No matter, the conclusion is that just as whether you are thinking something true or something false is constituted by a relation between you and the world, so also *what* you are thinking is constituted by a relation between you and the world.

These axioms and theorems are the rock-solid core of contemporary philosophy of mind, and if you doubt them (or something very much like them) you simply aren't in the game. Many of the arguments in Kim's textbook, and in many others, use them, and if they are abandoned so is the content of contemporary MMPM. The famous Chinese Room argument, for example, doesn't start, for its logical form is: For no A does A necessitate B because for any A Searle can imagine that A is true and B is not, and hence (by the Cartesian Axiom) it is possible that A and not B, where A is "executes a program" and B is "understands Chinese."

**4. Davidsonia.** The Identity Strategy has the most texture in contemporary MMPM because of the efforts of Donald Davidson, who proposed a version of the strategy almost everyone feels obliged to address, whether to endorse or, with Kim, to dispute. Davidson's essays on actions and events began with arguments about logical form, like nothing so much as the early work of Bertrand Russell. Besides disputing

3. I recognize the difference between being right or wrong because my keys are or are not in the car and being right or wrong because arthritis is or is not a disease of the muscles. In the first case I am wrong because of a particular fact, in the second case because of a generic fact. In both cases the truth or falsity of the utterances (respectively "My keys are in the car" and "I have arthritis in my thigh") depend as well on a linguistic practice, although the role of language in determining the truth or falsity of the corresponding *thoughts* is less clear. But I do not see that the second case should mark a difference of meaning and the first not, or what that difference is. In any case, I disdain MMPM not for the Putnam-Burge Theorem, but for the uses to which it has been put, as described in the later discussion of the "Extrinsicity of the Mental."



the separation of action from causation (physical and otherwise) advocated by Davidson's philosophical contemporaries of the time—Melden and Hampshire, for example—Davidson's analyses led him to the conviction that making logical sense of our talk requires that we refer to and quantify over a special class of particular entities, *events*.

Davidson's events are the entities that enter into causal relations. They have spacetime location, but they need not be all of what goes on in a spacetime region: distinct events can be in the same place and time. It is not clear whether each event must within itself have variations of features, or whether an event must differ in features (other than in spacetime location) from each of its causes. In the end, other than by giving examples of everyday descriptions, Davidson is able to say positively only that no two events have the same set of events as causes and the same set of events as effects.

A cause, according to Davidson, is an event that has features necessary and sufficient for features of some later event according to some causal law. A causal law is a true statement of a particular logical form and content, specifically:

$$[(e)(n)((Fe \ \& \ t(e) = n) \rightarrow (E!f)(Gf \ \& \ t(f) = n + \varepsilon \ \& \ C(e,f)) \ \& \\ [(e)(n) ((Ge \ \& \ t(e) = n + \varepsilon \rightarrow (E!f)(Ff \ \& \ t(f) = n \ \& \ C(f,e)))]]$$

where "the variables 'e' and 'f' range over events, 'n' ranges over numbers, F and G are properties of events, 'C(e,f)' is read 'e causes f,' and 't' is a function that assigns a number to an event to mark the time the event occurs." (Davidson 1980, 158)

Davidson says in the text that causal laws must be deterministic, as the formulation implies, but he takes the requirement back in a footnote, without further explanation. There are formal consequences: causation is irreflexive, and is not necessarily transitive.

Davidson writes of events as causes but features of events (or events as described in one way or another) as "necessary and sufficient as causes" (172). The phrasing is perplexing, but I take it that the meaning is not that the properties or features are *causes*, but rather that in the necessary and sufficient conditions in some causal laws. such properties or features are ascribed to causal events.

Causal laws, Davidson claims, hold only in physics, and perhaps also in essentially physical subjects such as parts of chemistry. They are to be distinguished from the sorts of pseudo-causal regularities that occur in psychology, the social sciences, perhaps even parts of biology. Generics (e.g., "Whales give milk") don't have the requisite logical form, nor do *ceteris paribus* claims ("Other things equal, metals expand when heated"). Psychological attributes depend on a web of other psychological attributes—no one has a belief, or desire, except as he also

has other beliefs and desires, and the varieties of such conditions are endless. So there are no psychological causal laws, and no psychophysical causal laws.

It seems of little importance in the literature of MMPM that Davidson gave not a single example of a causal law in physics nor any survey of psychophysics or neuropsychology to establish the absence of causal laws in those subjects. Davidson is almost certainly wrong on both counts: there are no causal laws of his kind in physics, but so nearly as there are, comparable generalizations contain psychological predicates. Davidson's physics is not as fantastic as Twin Earth, but fantastic nonetheless. Consider a real, if false, physics, simple dynamics. From the law of the pendulum and some initial parameters (the length of the pendulum arm, for instance) an uncountable infinity of propositions follow that may appear to be of the form Davidson requires of causal laws: The bob is at angle  $\theta$  with momentum  $\mu$  at time  $t$  if and only if it is at angle  $\theta'$  with momentum  $\mu'$  at later time  $t'$ . Substitute numbers for the parameters, take the states to be features of events and say the relation between the events are of cause to effect, and you have something of the form of one of Davidson's causal laws. Except you don't, because the "Law of the Pendulum" isn't a law of physics in any unconditioned sense. It assumes an endless array of conditions: Gravity does not change in the meanwhile (but of course it does), the length of the pendulum rod remains constant, there is no friction, nothing interferes with the motion of the pendulum. In classical physics nothing in the state of the universe at a time is inconsistent with the appearance at a later time of an object decelerating in from infinity to bollix up the pendulum, or any other system (see Earman 1988). The notion of a sufficiently "isolated" system is essentially contextual, and cannot be made both precise and general without falsifying the antecedents of "causal laws." Quantum mechanics presents other difficulties: it has a dynamical equation that relates states at one time to states at another, but the relations have a notorious *ceteris paribus* clause: that no observation that changes the state occurs in between times. The notion of an "observation" appears psychological, and might be removed by using instead some more general notion of "interaction." But there is no physical description of the isolation that prohibits interaction, and the moral of the Bell inequalities may be that there cannot be such a description (see Maudlin 1994). Davidson is optimistic:

Within the physical sciences we do find . . . generalizations . . . such that if the evidence supports them, we then have reason to believe they may be sharpened indefinitely by drawing upon future physi-

cal concepts: there is a theoretical asymptote of perfect coherence with all the evidence, perfect predictability (under the terms of the system), total explanation (again under the terms of the system). (219)

I don't know quite what this means, but insofar as I understand the passage, it seems likelier to be true of psychophysics and neuropsychology than of the causal claims of physics implied by fundamental differential equations. It may well be that the ordinary beliefs and desires we ascribe to one another have no necessary and sufficient physical conditions, but it seems very likely that, in humans, simple perceptions, pains, sexual arousal, and many other mental conditions do so.

The concluding pieces in Davidson's metaphysics are these: an event is mental if it has a mental description, physical if it has a physical description. So one and the same event can be both mental and physical. Davidson claims that all mental events are in fact physical events, although nothing in the story so far requires this (and it is hard to see why it should be believed—might not a mental event just be part of a physical event, or perhaps take up all of two or more physical events, or perhaps be parts of several distinct physical events?). But to continue: a mental event *c* can be a cause of another event *e*, whether physical or mental, because *c* instantiates the antecedent and *e* the consequent of a causal law, which is of course a physical law. The Mental Causation challenge is met.

**5. Kim's View.** So, my scientific cavils aside, what's not to like? Kim's formulation of the Mental Causation challenge goes this way: "*Given that every physical event that has a cause has a physical cause, how is a mental cause also possible?*" (38). Davidson's answer is that the mental cause and the physical cause are the same event. What Kim finds unsatisfactory in this answer is (1) that supervenience has no explanation, and (2) that mental properties have no causal role. The second difficulty is exacerbated by the Putnam-Burge Theorem, or what Kim calls the Extrinsicity of the Mental:

if mental activities are computational processes on beliefs, desires and such, it would seem that it is the syntactic shapes of these states, not their representational contents, that are causally relevant.

. . . the crux of the problem lies in the supposed fact that mental properties, in particular, content properties (e.g., being a belief that *P*), are relational properties, extrinsic to the organisms instantiating them, whereas we expect the causative properties of behavior to be intrinsic and internal. (35–37)

Kim essentially drops the problem, without an answer I can find. Let's try an analogy from *Beginning Electricity and Magnetism*: I have an arrangement of charges fixed in position in a region with constant gravitational field. If in the same region I release a test particle of unknown charge and mass, how it then moves depends on the unknown charge and mass and on where I release it. Some different arrangement of fixed charges might bring about the same motion of my test particle if it were released in the same place: only the gravitational field and the total electromagnetic field in the region in which I release the test particle matters to how the particle moves. In the two circumstances the particle has different relations with the field sources, the arrangement of the field sources is a cause of the test particle's motion, but the test particle's motion is independent of the arrangement of the field sources conditional on the electromagnetic field values in the region where the particle is released. There is nothing puzzling about this; we are not stymied by the question: "But then how does the *relation* between the test particle and the arrangement of the source charges cause the motion of the test particle." That question, so far as it has an answer, has been answered or obviated, and we are able to use the evidence of motion to infer something about the ratio of the particle's charge to its mass. Why is it any different with the Extrinsicity of the Mental? For similar reasons, I see no scientific point to refutations of "individualism" in psychology that insist, correctly enough, that the language of psychology is not just about computation but also about relations, including intentional relations (compare Burge 1986). Indeed it is; but cognitive psychology is only possible because psychologists can and do know the state of the external relata—the external world and circumstances, including the language they and their subjects speak. Condition on the values of the external relata and what remains to be studied is what, if anything, is "individualistic."

Kim's take on the problem of mental causation is that mental causation is not to be disputed, and the philosophical problem is to explain how it is possible, given that every physical event that has a cause has a physical cause; moreover any acceptable explanation must explain how supervenience comes about (or if you prefer, why it is the case). His principal concern is overdetermination: if mental events cause physical actions, and physical events also cause them, and the mental causes cannot possibly do their work unless some physical causes do theirs, how can the mental causes really be causes? Kim's answer is to reject Davidson's "anomalous monism" and to substitute an identification of mental states with functional roles that are realized by various physical systems with various properties. (Kim gives scarce consideration to one possible answer: the occurrence of instances of physical

properties are insufficient without the co-occurrence of distinct mental features—in other words, the true causal laws of Davidson’s form predicate mental properties of the causal events. That answer violates another premise of the literature on mental causation, namely that the physical world is “causally closed.”) The argument, both positive and negative, is through a tour of others’ views, and that is one of the many virtues of the book, but a virtue that makes it difficult to summarize. I will comment only on parts I found particularly interesting.

Kim considers the idea, attributed to Ned Block, that mental properties are “second order.” A second order property is the property of having some one (or perhaps several?) of a set of other properties. Examples are “provocative,” said, for example, of a red cape waved before a bull, and “primary color.” Block claims second order properties are “epiphenomenal”: they have no causal powers, which are vested instead in whichever first order property may instantiate a second order property in any particular context. Kim thinks otherwise:

If a second-order property F is realized on a given occasion by a first-order property H . . . then the causal powers of this particular instance of F are identical with (or are a subset of) the causal powers of H. (54)

Kim is free to talk as he wishes, but the how people who are not philosophers talk about causation when various levels of generalization are available seems likely to be intricate. The psychological evidence is fairly strong that in ascribing causal powers, especially in novel contexts, people spontaneously choose the most specific available property that differentiates cases in which a kind of effect occurs from those in which it does not (see Lien and Cheng i.p.). Give adults data on properties such as transparent, white, colored, and blue light and lenses, and ask them why a particular blue lens only transmits blue light, and they will usually tell you because its blue, not because its colored, and with similar experience with novel predicates, they will use that causal knowledge in predicting the novel properties of new cases for which the data given them is otherwise insufficient. I imagine background beliefs might bias such judgments in some contexts, and I suspect we are content to treat a more general feature as a cause of a kind of effect when its various instantiations all produce effects of that kind.

Even with his way of talking about the causal powers of “second order properties,” Kim does not think them sufficient to solve the Mental Causation challenge, because mental properties get instantiated on any occasion only by physical properties. (Kim makes essentially the same argument in several places against the sufficiency of appeal to different “levels” of causal explanation.) What more is required is some

kind of identification, an identification which he thinks is all but irresistible if the causal power of a mental property on any occasion just is the causal power of whatever physical property instantiates it on that occasion.

Kim considers several versions of the Humpty Dumpty strategy. One is Tyler Burge's. I reproduce part of Kim's quotation from an essay by Burge:

It would be perverse to think that the mentalistic explanation excludes or interferes with non-intentional explanation of the physical movement. I think that these ideas seem perverse . . . because we know that the two causal explanations are explaining the same physical effect as the outcome of two very different patterns of events. The explanations of these patterns answer two very different types of inquiry. Neither type of explanation makes essential, specific assumptions about the other. (Burge 1993).

Name-calling aside ("perverse" is used four times in the full passage Kim quotes), Burge may have something in mind analogous to the relationship of a biochemical explanation of a feature of an animal, on the one side, and an evolutionary explanation on the other. Kim's objection, which seems entirely correct to me, is that as one learns more, two explanations that make no contrary "essential, specific assumptions" may nonetheless come to bear evidence against one another. So biochemical discoveries can undermine evolutionary explanations, and vice versa. Kim thinks that once the Scientific Viewpoint is accepted, that is enough for physical and psychological explanations of action to stand witness against one another.

Kim discusses a proposal by Frank Jackson and Philip Pettit that draws an interesting analogy between explanations in terms of macroscopic, dispositional properties, as in explaining the breaking of a vase by its fragility, and explanations of a computer's input/output relations that specify a program in a high-level programming language. The fragility may not be a cause of the breakage but it is *causally relevant* because it asserts the existence of a property or properties that are causes; the program descriptions are not of causes but of circumstances that will be true of the machine in between input and output, even though underlying processes do all the causation. Kim thinks the proposal gives up on the Mental Causation challenge and is really a form of epiphenomenalism.

Kim's own solution to the Mental Causation challenge is this: mental properties are second-order properties, in roughly the sense described by Block, but as interpreted by Kim. A mental property is—*is* identically—a "causal role," which may be fulfilled (or instantiated or

realized, take your pick) by a variety of different physical properties. The causal power of a mental property of a particular system on a particular occasion is the causal power of whatever physical properties fulfill that causal role in that system on that occasion.

What's a "causal role"? We understand a role for actors—many different Hamlets can play the same role; the play is the same though the performance is not. Kim gives two less distant examples, borrowed from Kripke. What plays the causal role of producing heat can be many substances in many kinds of motion; what causes variations in its intensity is the mean kinetic energy of the motions, whatever their kind and substance. What plays the causal role of transmitting genetic information is DNA, although Kim supposes that other substances could do so. Unfortunately the examples don't quite meet the need. In each of the cases, temperature and genetics, a kind of effect is specified and understood, and the causal role is specified as the role of causing that *kind* of effect. If we understand "cause" and we understand the kind of effect, we understand the causal role, and it becomes an empirical question to discover what fulfills the role in various systems and circumstances. But with mental properties, no kind of effect is specified: we are simply told that mental properties are identical with *some* causal role, we know not which. What is the causal role of the belief that the Earth orbits the Sun, or of the belief that Clark Glymour is a grumpy guy? I don't think one can give an answer. It is ordinary enough to assign particular effects on particular occasions to such beliefs, but not generic, system-independent, occasion-independent causal roles. We may explain why Shlomo avoids Glymour at meetings by citing a belief and desire as the cause—Shlomo believes that Glymour is a grumpy guy and wants to avoid grumpy guys—but the generic causal role of the belief is amorphous. What causes people to avoid Glymour at meetings? The list of reasons, each a distinct mental ascription, is endless. Even Shlomo, on another occasion, may avoid Glymour for a different reason. If we abandon reference to kinds of effects, we are left with a schema that cannot be coherently filled out, or is hopelessly circular and makes no reference to causal role: the belief that Glymour is a grumpy guy is identical on any occasion with whatever the physical realization on that occasion of the belief that Glymour is a grumpy guy . . . (how *does* one complete the schema?); or else: the belief that Glymour is a grumpy guy is identical on any occasion with whatever on that occasion physical realizes the belief that Glymour is a grumpy guy. I think we are left only with this: two mental properties are different if and only if they have different causal roles.

While there are no doubt other sources as well, the contemporary idea that mental states are causal roles, the functionalist idea, can be

traced to the proposal by Hilary Putnam that mental states are computational states, or features of computational states, in a definite computer program. Whatever runs that computer program has those mental states. The trouble is that there is no reason to believe that there is a universal program shared by all of us, (although it receives different inputs in our several life histories) and that whoever believes Glymour is a grumpy guy (and I know there are a lot of you out there) shares a particular computational feature describable in the Programming Language of the Mind, a feature not shared by any of those who have no such belief. Something like that may actually be so for certain special mental states associated with perception, arousal, automatic motor control, and so on, but for the great variety of ordinary mental descriptions, there is not a whit of empirical evidence for the proposition. I don't mean to suggest that we cannot be truly and usefully described as biological computers. I do mean that while there may be some general psycho-computational identities, we have no empirical evidence to think there are enough of them to make Kim's proposal true, at least not for our ordinary mental explanations, the ones whose causal significance Kim aims to save.

Putnam's theorem, that mental states are "multiply realizable" by systems of different material kinds, is the foundation of almost all functionalist accounts, including Kim's. Arguments for the theorem inevitably appeal to instances of the Cartesian Axiom. The recognition of natural laws usually stymies such arguments: no one would argue that water can possibly be made of iron and oxygen simply because one can *imagine* iron oxide looking and tasting and feeling and satisfying like water. But no known natural laws prevent the creation of an android from non-biological materials, and so philosophers take license to have imagination imply possibility. But for all I think the philosophers know, mental life with non-biological material substrates may be like the impossibility of perpetual motion or the impossibility of accelerating a mass to a velocity greater than light: there is no *particular mechanism* that prevents the thing, but it just never works out.<sup>4</sup>

**6. Science and Mental Causation.** Two questions remain: Why do we think mental states cause actions; and do they? Different analogies with physics yield different intuitions. First we note that of itself, the scientific revelation of physical mechanisms that mediate belief and desire does nothing, and should do nothing, to undermine the conviction that mental states cause actions. We do not think the articulation of a mech-

4. I impute some such view to Searle, and, despite the reliance of his Chinese Room argument on the Cartesian axiom, for all I know his conclusion may be right.



anism or material of causation undermines a cause. We do not conclude that acid does not cause bicarbonate of soda to release a gas because we know the molecular properties that acids share and the molecular composition of bicarbonate of soda and the chemical reactions that occur when acid and soda are combined, their varying thermodynamics and rates, and the phase properties of their products. But, second, intermediate states are not always, or even usually, intermediate causes. Dissolving copper metal in sulfuric acid produces a green solution, which can then be neutralized and reduced to produce a crystalline powder that melts at a high temperature. We do not think the color of the liquid causes the melting point of the precipitate.

Our understanding of causation derives from our experience with the production of effects by the manipulation of causes, either by ourselves, or by others, or by circumstances, and from the scientific extrapolation of relations of dependence to systems we cannot manipulate and even to those Nature does not manipulate. We cause others to do things by causing them to believe and desire things: When I ask my compliant child (now *there's* a philosophical fiction) to shut the door, I cause in her the belief that the door is open and the desire to shut it, which together cause her to shut the door. On first pass, nothing is more natural, or more scientific, than to suppose that the belief we bring about causes the action that follows it. We notice beliefs and desires in ourselves, and the actions that follow them, and think the beliefs and desires cause the actions.

On further reflection, the evidence underdetermines the character of the process. It may be that my request causes both my daughter's belief and her action, but the belief does not cause the action. How is one to know? The conviction that conscious mental states cause actions would be undermined by evidence that physical mechanisms cause conscious mental states and physical mechanisms cause action, and the mechanisms are separable, so that with appropriate interventions the actions may be caused without the occurrence of the mental states. There is accumulating evidence, still insufficient, that the mechanisms that produce conscious mental states may indeed be separable from those that produce the actions we customarily attribute to conscious thought. There is, in the first instance, evidence that physiological processes from which voluntary actions may be predicted precede any experience of volition or decision by some 300 milliseconds or more. Of course, this is consistent with the mental state being an intermediate cause between the predictive physiology and the action. Automatism of various kinds are known, and in a sense familiar to every one of us. Cognitive neuropsychology continues to uncover cases in which action occurs without any sense of will or volition, most famously cases of alien hand move-

ments. Blindsight and many examples of implicit learning show cognitive processes protected from consciousness. I do not know if there are, among the depressing cases of cognitive neuroscience, people whose bodies carry out coordinated action of the kind that would ordinarily be thought to require planning, without any sense of awareness or volition, or if not yet, whether such cases will eventually be found.<sup>5</sup> But whether the mechanisms underlying conscious mental states are separable from the mechanisms of purposive action is what matters to our conception of ourselves, and that is an empirical, not a metaphysical, question.

For the kind of thing it is, Kim's book is wonderful. But it is interior redecoration of the house of MMPM, rearranging the pieces, throwing some out, adding some new. Meanwhile, the foundation is adrift. Better to build on science.

## REFERENCES

- Burge, Tyler (1986), "Individualism and Psychology", *The Philosophical Review* XCV: 3–45.  
 ———. (1993), "Mental Causation and Explanatory Practice", in J. Heil and A. Mele (eds.), *Mental Causation*. Oxford: Clarendon.  
 Davidson, Donald (1980), *Essays on Actions and Events*. London: Oxford University Press.  
 Earman, John (1988), *A Primer on Determinism*. Dordrecht: Kluwer.  
 Heil, John (1998), *Philosophy of Mind*. London: Routledge.  
 Kim, Jaegwon (1996), *Philosophy of Mind*. Boulder, CO: Westview Press.  
 ———. (1998), *Mind in A Physical World*. Cambridge, MA: MIT Press.  
 Lien, Yunnwen and P. Cheng (in press), "Distinguishing Genuine from Spurious Causes: A Coherence Hypothesis", *Cognitive Psychology*.  
 Maudlin, Tim (1994), *Quantum Non-Locality and Relativity*. London: Blackwell.  
 Spence, Stephen (1996), "Free Will in the Light of Neuropsychiatry", *Philosophy, Psychiatry and Psychology* 3: 75–90.

5. For a useful survey, see Spence 1996.