

# Metaphysics and Mental Causation

LYNNE RUDDER BAKER

DOES what we think ever affect what we do? The answer to this question may seem obvious: of course, having the thoughts that we have leads us to do what we do. But philosophers have recently cast doubt on the relevance of what we think to what we do. Mental causation—the causation of what we do by what we think—becomes a problem against a particular metaphysical background. My aim is twofold: first, to root out the metaphysical assumptions that generate the problem of mental causation and to show that they preclude its solution; second, to dissolve the problem of mental causation by motivating rejection of one of the metaphysical assumptions that give rise to it.

There are three features of this metaphysical background picture that are important for our purposes. The first concerns the nature of reality: all reality depends on physical reality, where physical reality consists of a network of events.<sup>1</sup> The second concerns the nature of causation: causation is conceived of as 'objective', a 'real relation' in nature, instances of which are independent of anyone's explanatory interests.<sup>2</sup> The third concerns the conception of behaviour: behavi-

I am grateful to Derk Pereboom and Hilary Kornblith for helpful conversations and critical comments. Pereboom gave patient advice on multiple drafts.

<sup>1</sup> Terminology is a problem in the literature. Although the term 'event' is prominent in the writings of both Donald Davidson and Jaegwon Kim, the two philosophers have in mind different things. Davidson takes an event to be a concrete particular, with a particular spatio-temporal location. Kim takes an event to be an instantiation of a property at a time. When discussing Davidson, however, Kim sometimes engages in what seems to be a hybrid use: e.g. Kim charges that on Davidson's anomalous monism, 'an event's mental properties make no causal difference' (Kim 1989*b*, p. 35). I shall follow Kim in this hybrid use. A Davidsonian may replace talk of what properties an event has by talk of what descriptions an event has. A Kimian may replace talk of events as items having properties by talk of objects.

<sup>2</sup> On such a view, 'every event has a unique and determinate causal history whose character is entirely independent of our representation of it' (Kim 1988*a*, p. 230).

our is to be understood in terms of events with certain kinds of internal causes.

The internal events that cause behavioural events (are thought to) have various kinds of properties, some physical and some intentional. The intentional (or mental) properties are what I shall call 'content-properties': a desire to get a beer from the refrigerator, a belief that Summer Squall will win the Kentucky Derby. As I am using the term, 'content-properties' are e.g. those properties in virtue of which beliefs are themselves true or false.<sup>3</sup> When we explain behaviour in terms of a person's reasons, what we cite in explanations are various content-properties—properties determined by what she believes, what she desires, and so on. In this picture, these content-properties are attributed to the person's internal events that cause the behaviour.

In the wake of these assumptions, suspicion about mental causation comes naturally: assuming that a bit of behaviour is an event caused by internal events that have content-properties (such as a belief that you left your keys on the counter), it may well be that content-properties of the cause are causally inert. To borrow an example from Fred Dretske (1989), the soprano may be making meaningful sounds when she hits the glass-shattering high C, but the meaning is irrelevant to the properties (e.g. amplitude) of the sound that causes the glass to shatter. The fear concerning mental causation is that all content-properties may be like those of the soprano's high C.

Understood against this background, the question with which we started, 'how can what we think affect what we do?', is recast as this question: 'how can content-properties of internal events be causally relevant to producing behavioural events?' The problem of mental causation is to answer this question.<sup>4</sup>

A widespread assumption about the nature of explanation produces a corollary to the problem of mental causation. The thesis about explanation is that an explanation of a behavioural event

<sup>3</sup> I am not concerned here with some alleged narrow component of beliefs.

<sup>4</sup> Jaegwon Kim gives an account of mental causation in terms of the intrinsic properties of the agent. Since, arguably, content-properties depend in part on the agent's embedding in an environment, content-properties seem not to be intrinsic but relational. Hence, Kim declines to include content-properties in his account—an account which he restricts to properties like having a pain: having a pain may be a supervenient cause of quickly removing one's hand from a fire. As we shall see, however, the distinction between intrinsic and relational properties is less clear than it may at first seem.

mentions a causally relevant property of an internal event that causes the behaviour. Given this thesis and the background picture, if content-properties are causally irrelevant, then they are also explanatorily irrelevant. In that case, the shattering of the glass cannot be explained by what the high C meant. Moreover, if content-properties turned out to be causally, and hence explanatorily, irrelevant to behaviour, then (again, in light of the background picture) we should conclude that reasons, identified by content-properties, have never explained anything that anyone has ever done. It would not be just that putative explanations by reasons turned out to be second-class explanations; rather, they would not be explanatory at all.

What I want to show is that, given this metaphysical picture, the problem of mental causation is insoluble.<sup>5</sup> We simply have no answer to the question 'How can mental events, in virtue of having mental properties, make a difference to behaviour?' because the very assumptions that generate the question render it unanswerable. Moreover, I want to show that the metaphysical assumptions with which we began inevitably lead to scepticism not only about the efficacy of contentful thought, but about macro-causation generally. But if we lack warrant for claiming that macro-properties are generally causally relevant, and if we take explanations to mention causes, then most, if not all, of the putative explanations that are routinely offered and accepted in science and everyday life are not explanatory at all. So, I shall argue, we have an impasse: we must either give up (part of) the metaphysical background picture or give up almost all explanations that have ever been offered for anything. Since the generalization of the problem to macro-causation depends only on the theses about reality and causation, I shall here neither examine the conception of behaviour as events with internal causes nor the conception of beliefs as internal states.<sup>6</sup> Rather, I shall focus on the theses about reality and causation, and suggest that we stick with the explanations that have proved their worth and let the metaphysical chips fall where they may.<sup>7</sup>

<sup>5</sup> Kim (1989a, 1989b) goes some distance in this direction. Although I want to extend his arguments, my motivation is somewhat different from his: whereas Kim finds the background picture attractive, I do not.

<sup>6</sup> I consider these conceptions of behaviour and of belief in 'What Beliefs Are Not' (in preparation).

<sup>7</sup> NB, the metaphysical chips will not fall into two neat piles—mental and physical.

## 1. The Metaphysical Background

There are two obstacles to an affirmative answer to the question 'Is the fact that an event has a content-property causally relevant to behaviour?' First, since content-properties are arguably relational properties that depend on an agent's embedding in an environment, the proponent of mental causation must show how relational properties can be causally relevant to behaviour.<sup>8</sup> Since this issue is already under discussion,<sup>9</sup> and since the other obstacle to solving the problem of mental causation seems to me both more fundamental and more intractable, I shall say nothing more about providing a causal role for relational properties. Rather, I shall focus on the second obstacle: namely, that the metaphysical background picture itself contains elements that call into question mental causation.<sup>10</sup>

Two metaphysical theses, both of which are parts of the background picture, generate the problem of mental causation. The first is the thesis of materialism (roughly, the thesis that every property-instantiation supervenes on physical property-instantiations), and the second is the thesis of the 'causal closure of the physical' (roughly, the thesis that every physical property-instantiation that has a cause at *t* has a complete physical cause at *t*). Let us consider the import of these theses in order to formulate them more precisely.

On the intended interpretation, the thesis of the causal closure of the physical entails that 'if we trace the causal ancestry of a physical event, we need never go outside the physical domain'.<sup>11</sup> I take this to mean that the occurrence of the event, and its having a particular physical property, both have sufficient causes wholly in the physical domain. Thus, for any event that has a physical property—whether

<sup>8</sup> For arguments that content-properties depend on environment, see Burge 1979, Van Gulick 1989, and Baker 1987.

<sup>9</sup> For example, see Dretske 1988 and Millikan 1989g. At least some relational views of content-properties have internal difficulties, I believe. See Baker 1989, 1991a, 1991b.

<sup>10</sup> Philosophers who are concerned with similar metaphysical issues include McLaughlin 1989, and LePore and Loewer 1989. Fodor tries to bypass these issues (Fodor 1989).

<sup>11</sup> Kim 1989b, p. 43. The discussion of non-reductive materialism borrows heavily from Kim.

or not it has mental or other properties—there are sufficient physical conditions for its occurrence and for its having all of its physical properties.

Now, what is a physical property? Often, philosophers speak as if neurophysiological properties were physical properties. But are they physical in the sense needed for the causal closure of the physical? No. If we interpreted the causal closure of the physical to imply that every neurophysiological event that has a cause at  $t$  has a complete neurophysiological physical cause at  $t$ , then the thesis would be obviously false. For not every neurophysiological event at  $t$  has a complete neurophysiological cause at  $t$ : neurophysiological laws will themselves be hedged by *ceteris paribus* clauses. 'Lower-level' phenomena at the molecular or even quantum level will interrupt neurophysiological processes that are otherwise governed by neurophysiological laws (cf. Fodor 1981*b*). So, there is no causal closure at the neurophysiological level. A system is causally closed if and only if the elements of the system interact causally *only* with other elements of the system; there is no causal influence from 'outside' the system. Since the notion of causal closure of the physical applies only at the lowest level of micro-physics, the only sense of 'physical' in which the causal closure of the physical has a chance of being true is micro-physical, where 'micro-physical' is a name for whatever turn out to be basic physical particles and their properties. So, let us formulate the thesis of the causal closure of the physical like this:

CCP Every instantiation of a micro-physical property that has a cause at  $t$  has a complete micro-physical cause at  $t$ .

Now let us turn to the thesis of materialism: every instantiation of a property supervenes on instantiations of micro-physical properties. How are we to understand the relation of supervenience? First and foremost, the supervenience relation is to be a relation of dependence or determination. On the intended interpretation, materialism is supposed to capture the idea that everything that happens depends on, or is determined by, what happens at the micro-physical level. Some philosophers have tried to formulate the relevant sense of supervenience as 'global supervenience':

GS Two possible worlds that are indiscernible in all micro-physical respects are indiscernible in all respects.<sup>12</sup>

<sup>12</sup> The usual formulations of global supervenience refer to physical, not to

But GS alone does not entail that the micro-physical determines the mental and hence by itself is insufficient to support materialism.<sup>13</sup> For it entails nothing about the ways that properties are distributed over individuals within possible worlds. It is consistent with GS, for example, that there be two individuals in physically similar environments within a given world who have all the same physical properties, but who differ mentally. For GS, supplemented by connective laws, is silent about the relation between micro-physical and higher-level (e.g. mental) properties within a given world. GS is not violated as long as all other physically indiscernible possible worlds also have two individuals who are physically alike but mentally different.<sup>14</sup>

Thus, GS does not even imply that an individual's mental properties (relational or not)<sup>15</sup> are determined by her micro-physical properties. Since what is needed for materialism is a stronger notion of supervenience, I shall adapt Kim's construal of 'strong supervenience' for the purpose of explicating Materialism:<sup>16</sup>

SS Necessarily, for any instantiation of any property, *F*, there is an instantiation of a micro-physical property, *G*, and, necessarily, anything that has *G* has *F*.

Although I believe that SS captures the basic idea that everything that happens is determined by what happens at the micro-physical micro-physical, properties. But I formulate GS with reference to micro-physical properties to make GS parallel to CCP. Anyway, I believe that, for purposes of materialism, it is assumed that two worlds are physically indiscernible if and only if they are micro-physically indiscernible. For (however one construes 'physically') if two physically indiscernible worlds were micro-physically discernible, then the physical world would not encompass the micro-physical; and if two micro-physically indiscernible worlds were physically discernible, then the physical would be determined by something other than the basic micro-physical elements—i.e. the materialism would not 'go all the way down'.

<sup>13</sup> This point is made forcefully in Kim 1989b.

<sup>14</sup> John Post has constructed a model to show that supervenience does not entail the existence of connective laws. Granted, but without connective laws, supervenience is compatible with the falsity of 'all truth is determined by physical truth', as we have just seen. See Post 1990, 1987. Petrie (1987) has a similar example, which is criticized by John Heil 1992.

<sup>15</sup> The point applies whether or not mental properties are taken to be relational, as long as relational properties are construed in such a way that (e.g.) 'being in a world with *n* electrons' does not count as a relational property in the relevant sense.

<sup>16</sup> Kim 1984a. Kim defines strong supervenience as an abstract relation between families of properties, but it is unclear to me that all properties are naturally sorted out into families. (See my comment about 'levels' below.) For my purposes, I want to formulate strong supervenience as a substantive thesis about the nature of reality.

level, SS still requires some elaboration. For even if content-properties ultimately depend on micro-physical properties, they do not directly depend on micro-physical properties. Presumably, content-properties are determined by neurophysiological properties and the relations of the neurophysiological properties to other (e.g. environmental) properties, and neurophysiological properties depend on molecular biological properties, which in turn depend on chemical or physical properties, and so on down to the micro-physical level. The 'levels' metaphor makes most sense in scientific contexts. For those properties for which the 'levels' metaphor is apt, however, let us replace SS with a definition that allows for intermediate 'levels':

SS\* Necessarily, for any instantiation of a property  $F$  at level  $i$  ( $i = 1, \dots, n$ ), there is an instantiation of a property  $G$  at level  $j$  ( $0 \leq j < i$ ), and necessarily, anything that has  $G$  has  $F$ ; and all properties at level 0 are micro-physical properties.<sup>17</sup>

Since I do not think that we have a well-defined idea of 'levels' that is sufficiently general to accommodate the supervenience of all properties (e.g. the property of 'being three feet from an ocean', or of 'being a bachelor'), I prefer SS, and introduce SS\* only as an elaboration.

Several features of the intended interpretation of SS deserve mention. The first four may be stated briefly. The fifth requires explanation.

1. SS is a metaphysical, not an epistemological thesis; we may never discover the relevant micro-physical properties on which higher-level properties supervene. So, SS does not imply reducibility in any epistemological sense.
2. SS does not require that  $G$  be a 'well-behaved' micro-physical property mentioned in the laws of fundamental physics.  $G$  may be hideously complex, a Boolean combination of properties that occur in the laws of fundamental physics (and their relations—see below). So, there is no assumption that the supervenience base  $G$  is a natural kind.
3. SS also allows for 'multiple realization' of higher-level properties: the property of being a desire to go to the beach, for example, may be realized in any number of different micro-physical properties (and their relations).

<sup>17</sup> The necessity required for SS\* need not be as strong as metaphysical necessity. For purposes here, 'necessarily' may be read as 'it is a law that'.

4. The bearers of the micro-properties are not the same as the bearers of the macro-properties: Jill has a pain; the particles in Jill's body have spin; but Jill doesn't have spin even if the pain ultimately supervenes on the spin of the particles.<sup>18</sup>
5. If the bearers of the micro-properties are micro-parts of, say, medium-sized objects, then, for supervenience to hold, the micro-parts' relational properties must be included in *G*. For example, consider a ball's property of being round. This property instantiation supervenes on the properties of the ball's micro-parts; but the micro-parts' intrinsic properties do not suffice as a supervenience base for the ball's roundness. For something having micro-parts with exactly the same intrinsic properties may fail to be round, because the micro-parts have different relations to each other. For suppose that there are two qualitatively similar round balls. The object that is the mereological sum of the upper hemisphere of one ball and the lower hemisphere of the other has micro-parts with the same intrinsic properties as either of the balls, but this new object is not round. Therefore, *G* is not only a perhaps infinitely complex property, but is itself composed of relational properties. The ball's intrinsic properties supervene on its parts' intrinsic and relational properties.

So, we must admit that the micro-properties in the supervenience base *G* are relational. But from the micro-perspective, we have no distinction relevant to supervenience between the relational properties whose relata are micro-particles of the same macro-object, and relational properties whose relata are micro-particles of different macro-objects. Since the supervenience base, *G*, includes micro-particles' relations to micro-particles of different macro-objects anyway, then the macro-property *F* that supervenes on *G* may itself be a relational property. The intrinsic-relational distinction at the macro-level is simply irrelevant to the metaphysics of supervenience. For example, the ball's property of 'being three feet from the basket' supervenes on the ball's micro-parts' properties and their relations as much as the ball's property of 'being round'. Indeed, according to general relativity theory, the property

<sup>18</sup> On Kim's formulation, *x*'s having *F* supervenes on *x*'s having *G*, where *F* may be a macro-property and *G* a micro-property. It would be more accurate to say that *x*'s having *F* supervenes on *x*'s having parts that have *G*. Kim takes up this issue in his 1988*b*.



expressed by 'x is three feet away from y' supervenes not only on the properties of the smallest spatio-temporal region that includes objects so related, but also on the distribution of matter in other parts of the universe. So, SS is comprehensive in the sense that it grounds all properties—any higher-level object's relational as well as intrinsic properties—in micro-physical properties, relational and non-relational.

In the light of these features, I believe that SS and SS\* provide the weakest notion of supervenience that can underwrite materialism.<sup>19</sup> So, I shall take the metaphysical background picture to require CCP and SS (or SS\* wherever the 'levels' metaphor is appropriate.)

## 2. The Problem of Mental Causation

To see how CCP and SS\* generate the problem of mental causation, suppose that, at a certain auction, Jill wanted to make a bid, moved her eyebrow in a certain way and thereby made a bid. According to CCP, there is a complete micro-physical cause of Jill's eyebrow's moving—say, a brain-event's having parts with micro-physical property *p*, where *p* may be a heterogeneous disjunction of properties and their relations. Let us suppose also, via SS\*, that there was a neurological event that was sufficient for her eyebrow's moving. Then, we have:

- (a) The event that caused Jill's eyebrow to move did so in virtue of having neurophysiological property *N*.

The question is whether or not Jill's wanting to make a bid is also causally relevant to her eyebrow's moving. Assuming that (a) is not only true, but also gives the complete cause of the eyebrow movement, then Jill's desire to make a bid is causally relevant to her eyebrow's moving just in case the following is also true:

- (b) The event that caused Jill's eyebrow to move did so in virtue of being a desire to make a bid.

Now, given that (a) is true, is there any reason to think that (b) is also true? We are assuming that the neurological event that sufficed for the eyebrow movement also had (or realized) the property of

<sup>19</sup> Even proponents of what is called 'non-reductive materialism' should endorse SS, if they are really materialists. For a defence of non-reductive materialism, see Pereboom and Kornblith 1991.

being a desire to make a bid. But, of course, it does not follow from its being a desire that it was causally relevant to the behaviour. As we have seen, not every property of an event is causally relevant to what the event causes. For suppose that later Jill gave an interview to the *New York Times*, which mentioned her desire to make a bid. As it happened, her desire to make the bid was the first event referred to on page 27 of the *Times* on 12 September 1990. Then, since her desire to make the bid had the property of being the first event referred to on page 27 in the *New York Times* on 12 September 1990, on Davidsonian views at least, the first event referred to on page 27 of the *Times* on 12 September 1990 caused Jill's moving of her eyebrow. But causal relevance concerns properties (or descriptions) of events, and no one would hold that being the first event referred to on page 27 was causally relevant to Jill's moving her eyebrow. (For one thing, the desire acquired the property of being the first event referred to on page 27 days after Jill made the bid.)<sup>20</sup> So, we cannot conclude that the desire was relevant to the eyebrow movement simply because the relevant neurophysiological event that produced the eyebrow movement had the property of being a desire.

So, what would entitle one to conclude that the desire to make a bid was causally relevant to the eyebrow movement? If there were a law that any instantiation of *N* is sufficient for a desire to make a bid, then it may be thought that Jill's desire to make a bid is causally relevant to her eyebrow movement. More generally, one could try to secure the causal relevance of the desire by invoking this principle:<sup>21</sup>

- P* If an instantiation of property *G* at *t* is a complete cause of an instantiation of *H* at *t'* (*t'* > *t*), and, necessarily, any instantiation of *G* at *t* is sufficient for an instantiation of *F* at *t*, then the instantiation of *F* is causally relevant to the instantiation of *H* at *t'*.

Using *P*, we may argue as follows: by assumption, the instantiation of property *N* is a complete cause of the eyebrow movement;

<sup>20</sup> This example is a variation on the kind of example that Davidson (1970) advanced.

<sup>21</sup> I am adding references to times here to avoid obvious counter-examples to *P* in which the instantiation of *G* is a common cause of two later events that instantiate *H* and *F*, respectively. The need for the reference to times was pointed out by Alexander Rosenberg.

so if, necessarily, all instantiations of *N* are desires to make a bid, then, by Principle *P*, the desire to make a bid is causally relevant to the eyebrow movement. Let us assume for the moment that, necessarily, all instantiations of *N* are desires to make a bid.<sup>22</sup> (Such an assumption is, in fact, false since someone could be in neurological state *N* in a world without auctions or conventions for bidding, in which case the instantiation of *N* would not be a desire to make a bid. But this problem is irrelevant to the illustrative purpose of the example.) So—it may be claimed—by *P*, Jill's desire to make a bid is causally relevant to her eyebrow's moving.

However, this argument is problematic because *P* is subject to counter-examples. According to *P*, properties that intuitively are not causally relevant get counted as causally relevant. Here are a couple of examples:

- (i) Suppose that a certain woman's having measles during her pregnancy was a complete cause of an instance of a birth defect, and suppose that, necessarily, anything that has measles at *t* has red spots at *t*. Then it would follow from *P* that having red spots was causally relevant to the baby's having a birth defect. In that case, it is in virtue of the mother's having red spots that the baby had a birth defect. But the only relevance of having red spots is that red spots are a symptom of measles; it is not in virtue of, or because of, the woman's having red spots that the baby had a birth defect. Symptoms are just the sorts of things that we want to rule out as having causal relevance to the effects of disease.
- (ii) Suppose that Moby Dick's being a large whale at *t* was the complete cause of Captain Ahab's interest in him at *t'*, and suppose that necessarily, anything that is a whale at *t* is a mammal at *t*. Then, it would follow from *P* that Moby Dick's being a mammal at *t* was causally relevant to Ahab's interest at *t'*. Then, it would be in virtue of Moby Dick's being a mammal that Moby Dick caused Ahab to become interested in him. But, intuitively (I should check with Melville about this), Moby Dick's being a mammal played no causal role in Ahab's interest.

These examples render *P* problematic and hence not available to secure the causal relevance of Jill's desire. Given CCP, SS\* and (a), we need some warrant for (b). Principle *P* looked promising, but seems subject to counter-examples. The fact (if it were a fact) that

<sup>22</sup> But not, of course, conversely, if we allow for multiple realization.

Jill's desire supervened on instantiations of physical properties would not be enough to show that the desire—in addition to the physical properties on which it supervened—was causally relevant to the eyebrow movement. In light of the background picture, then, we need an independent reason to think that content-properties are causally relevant.

I am not claiming that we always need a principle in order to make any distinction in a non-*ad hoc* way. We may have no line that determines when a fetus is a person, say; yet it is clear that a fertilized egg that is not yet attached to the uterine wall is not a person, and that an adult human being is a person. But there are at least two salient differences between the person case and the case of causally relevant properties, where we do require a principle. (i) There is a continuous process from the fertilized egg to the adult; but whether or not a property is causally relevant to behaviour has nothing to do with drawing a line in a continuous process. (ii) We have clear cases of adult human beings that are persons; the only problem in the person case is where to draw the line, not whether there is something on one side of it. But what is at issue in the case of the causal relevance of content-properties is not where to draw the line, but whether or not there is a line to be drawn. So, to claim that the failure of *P* calls into question the causal relevance of content-properties is not to suppose that we need some principle to warrant making every distinction.

To sum up the discussion so far: all significant versions of materialism (I think) are committed to CCP and SS; yet, given CCP and SS, there seems no way to avoid the unhappy conclusion that content-properties have never explained any behaviour, and hence that no explanation of behaviour in terms of an agent's reasons has ever been correct. We seem to be at an impasse. On the one hand, CCP and SS (or similar theses) are well-entrenched metaphysical views; on the other hand, they render the problem of mental causation insoluble.

### 3. From Mental Causation to Macro-causation

Before looking for ways out, let us consider the extent of our metaphysical predicament. The problem of mental causation is generated without any assumptions about intentionality, without any assumptions about the existence or non-existence of belief/

desire laws or of any psycho-physical laws, and without any assumptions about the nature of a correct theory of behaviour. All that is needed to generate the problem of mental causation are the two widely held metaphysical assumptions CCP and SS (or SS\*).

In fact, the difficulty is worse than it may appear, for the problem of mental causation generalizes to a comprehensive problem of macro-causation. The idea underlying CCP and SS is that what happens at the micro-level determines everything else that happens: every instantiation of every property supervenes on an instantiation of some micro-physical property, and every instantiation of every micro-physical property (that has a cause) has a complete micro-physical cause. The general problem is this: since every property of every event supervenes on a property that has a complete cause in terms of micro-physical properties, we need some warrant to hold that any macro-physical properties are causally relevant to anything. If every instantiation of every property has a complete micro-physical cause, what room is there for macro-causation?

One may try to avoid the sceptical conclusion by holding that the explanatory success of the special sciences is itself warrant to take macro-physical properties to be causally efficacious. In the context of the purely metaphysical argument given so far, however, appeal to the explanatory success of the special sciences is to no avail. For, on the metaphysical picture, causation is metaphysically prior to explanation. So when SS and CCP call into question the causal efficacy of macro-properties, they *ipso facto* call into question the explanatory success of explanations that mention macro-properties. Given SS and CCP, the apparent explanatory success of the special sciences may well be just that: apparent, not genuine. If one wants to retain SS and CCP without scepticism about macro-causation generally, then one needs a metaphysical account that shows how macro-causation is possible.

Jaegwon Kim has offered an account of macro-causation which, he says, respects the closed causal character of the physical. The idea is that macro-causation reduces to micro-causation. If  $F$  and  $G$  are macro-causal properties, the macro-causal relation of  $x$ 's having  $F$  and  $y$ 's having  $G$  is reduced as follows: ' $x$ 's having  $F$  supervenes on  $x$ 's having  $m(F)$ ,  $y$ 's having  $G$  supervenes on  $y$ 's having  $m(G)$ , where  $m(F)$  and  $m(G)$  are micro-properties relative to  $F$  and  $G$ , and there is an appropriate causal connection between  $x$ 's having  $m(F)$  and  $y$ 's having  $m(G)$ '.<sup>23</sup>

<sup>23</sup> Kim, 1984*b*, p. 262. According to Kim's schema, the same individual that has  $F$

Let me show why I believe that this account accords no causal efficacy to macro-properties. In light of CCP and SS, what we need is a warrant for taking macro-causation to be more than a 'mere chimera'. But what Kim has provided cannot be that warrant, because causally related micro-properties have the same relation to non-causally related macro-properties as to causally related macro-properties. So, postulating causally related micro-properties on which macro-properties supervene is insufficient to establish that the macro-properties are causally related. Therefore, Kim's picture cannot account for genuine macro-causation. In greater detail, if Kim's schema for reducing macro-causation is correct, then  $P'$  should be true:

$P'$  If an instantiation of a micro-physical property  $m(F)$  is a complete cause of an instantiation of micro-physical property  $m(G)$ , and necessarily, for every instantiation of  $m(F)$ , there is an instantiation of  $F$ , and necessarily, for every instantiation of  $m(G)$ , there is an instantiation of  $G$ , then the instantiation of  $F$  is causally relevant to the instantiation of  $G$ .

Now  $P'$  would allow instantiations of non-micro-physical properties to be causally relevant to instantiations of other non-micro-physical properties. But is  $P'$  acceptable? First, notice that all the causal work allowed by  $P'$  is done by instantiations of micro-physical properties. Non-micro-physical properties are deemed causally relevant by fiat. Second,  $P'$ , like  $P$ , seems susceptible to counter-examples.

Let  $m(F)$  be micro-physical properties of a certain sample of sulphur, and  $m(G)$  be micro-physical properties of the dying of a certain bush. Let  $F$  be an instantiation of the characteristic yellow colour of sulphur (where we assume colour to be a spectral reflectance property, not a property of experiences), and  $G$  be the death of a certain bush. Now suppose that the instantiation of  $m(F)$  was a complete cause of an instantiation of  $m(G)$ . Then, according to  $P'$ , the instantiation of the yellow colour caused the death of the bush. That just seems false: although the instantiation of the yellow colour is a major clue to the cause of the death of the bush, it is a clue only because it is a sign of the presence of sulphur. Intuitively, what caused the death of the bush was the presence of sulphur; the

also has  $m(F)$ . As noted above, however, the bearers of the micro-properties are not the same as the bearers of the macro-properties. This can be remedied by taking  $x$ 's having  $F$  to supervene on  $x$ 's having micro-parts that have  $m(F)$ .

instantiation of the yellow colour was causally inert with respect to the death of the bush.

Or consider a different kind of case: suppose that a person exercising in front of a mirror jumps up and comes back down. Let  $m(F)$  be micro-physical properties of the space-time region that includes the mirror and the exerciser as she goes up, and  $m(G)$  be micro-physical properties of the space-time region that includes the mirror and the exerciser as she comes down; let  $F$  be some macro-properties of the reflection in the mirror as the exerciser goes up, and  $G$  be some macro-properties of the reflection of the exerciser as she comes back down. Necessarily, anything that has  $m(F)$  has  $F$ , and necessarily, anything that has  $m(G)$  has  $G$ . Therefore, by  $P'$ , the instantiation of the properties of the reflection in the mirror as the exerciser went up caused the instantiation of the reflection's properties as she came down. But, again, intuitively, the reflection has no causal efficacy.

So, the same kinds of causal micro-processes underlie non-causal macro-processes as underlie causal macro-processes. Therefore, even assuming SS, the existence of causal micro-processes cannot provide the warrant to take macro-properties to be causally relevant to anything.

One may try to block the radically sceptical conclusion by claiming that Kim's account is not intended to allow us to infer anything about macro-causation from micro-causation; rather, the direction of fit may only go the other way. Given that we have a case of macro-causation, we can give an account of it in terms of micro-causation. However, since the problem was that CCP and SS seemed to leave no room for macro-causation, one needs a reason to believe that macro-causation exists. If the existence of micro-processes on which macro-processes supervene is compatible with the macro-processes' not being causal at all, then Kim's account is insufficient. It certainly cannot supply any warrant for taking macro-causation seriously in the face of CCP and SS.

If we need a warrant which we lack, many would counsel scepticism about the causal relevance of macro-physical properties generally. In that case, the implications for explanation are equally disastrous. For, as mentioned earlier, another background assumption is that correct explanations cite only causally efficacious properties. So, if macro-physical properties are never causally efficacious, then we also have no warrant to hold that any explanation in terms of macro-physical properties is correct.<sup>24</sup>

<sup>24</sup> In the 'one philosopher's *modus ponens* is another philosopher's *modus tollens*' vein, let me anticipate by suggesting that macro-properties generally (and content-

The conclusion that macro-properties lack causal efficacy is cognitively devastating. The same reasoning that precludes solution to the problem of mental causation leads to scepticism about explanations in the special sciences and much of physics. On the metaphysical picture, even 'the acidity of the liquid caused the litmus paper to turn pink' is unwarranted. The actual causal transaction—the 'real relation in nature'—took place at the micro-physical level. On Kim's account of macro-causation, the acidity of the liquid no more caused the litmus paper to turn pink than the smoke of the fire caused the ceiling to collapse.

Indeed, none of the paradigmatic core cases, which originally define our concept of cause, would turn out to be genuine cases of causation. (Perhaps I am not alone in considering this consequence a *reductio ad absurdum* of the metaphysical assumptions.) The irony is that philosophers take the metaphysical background picture from which CCP and SS are derived to be 'the scientific world-view'. On the contrary, the metaphysical picture seems to undermine most of science as well as almost all of common sense.

#### 4. A Modest Proposal

We have a choice: either we give up one or more of the metaphysical theses, or we give up (almost) all of our causal understanding of the way the world works. Many philosophers think that the metaphysical picture from which the theses stem has much to recommend it. Without emphasizing (or perhaps even recognizing) the consequences of theses like CCP and SS, they think that the picture is required by a rigorous scientific outlook. I hope to have shown that not only does the picture render unwarranted any attempt to explain what we do by what we think, not only does it subvert our ordinary causal notions that are constitutive of law, morality, and everyday life, but also that it makes a mockery of the causal claims and explanations in the special sciences.

These consequences seem to me sufficient motivation to give up at least part of the metaphysical picture. But what part? Although there are three aspects of the picture—conceptions of reality, properties specifically) are explanatory, and hence causal. To take this line is to abandon the metaphysical picture; for, as we have seen, we have no way to fit content causation and macro-causation into the metaphysical picture.



causation, and behaviour—that generate the problem of mental causation, I have focused, and shall continue to focus, only on the conceptions of reality and causation. It is not that I think that the conception of behaviour is innocent, but that the conceptions of reality and causation by themselves give rise to a comprehensive problem about macro-causation generally.

The trouble caused by the conjunction of SS and CCP can be avoided by rejecting CCP and rethinking the notion of causation. Although I believe SS to be logically consistent with my proposal, I think that it is a gratuitous bit of metaphysics whose role is primarily to fulfil a need for a totalizing thesis. I would simply point out that SS has less to do with science than with the central realist intuition, which I do not share. The intuition is that metaphysics can be done in total isolation from epistemology: we can say what, of metaphysical necessity, is the case without regard to how we might come to know it apart from metaphysical reasoning. It is doubtful that we will ever be able to ascertain the micro-physical supervenience base of macro-property instantiations generally. For example, the micro-physical properties in virtue of which the thing that I am sitting on is a chair are vast; they include not only the properties of the particles that compose the chair, but the properties of particles on which our conventions and intentions supervene. Something is a chair in virtue of the use to which it is put, the intentions of its maker, and so on, not in virtue of the properties of its constituent parts. Since I do not believe that we will ever be in a position to trace very many (if any) property instantiations to any micro-physical base, SS seems to me to be idle speculation.<sup>25</sup> So, let us leave SS alone and turn to CCP, which is the real culprit for causation.

CCP is founded on the idea that causation is an 'objective relation' between events in the following sense: 'that it is instantiated does not entail anything about the existence or non-existence of any intentional psychological state—in particular, an epistemological or doxastic state—except, of course, when it is instantiated by such states' (Kim 1988a). This seems to me an unacceptably narrow characterization of causation. On this view,

<sup>25</sup> Indeed, SS suggests an inverse of the Great Chain of Being. Whereas in the original Great Chain of Being, the 'higher' in the chain the greater the degree of reality; in the contemporary Great Chain of Being, the 'lower' in the chain, the greater the degree of reality.

'taking a deduction for the office in your home caused you to be audited by the IRS', would seem to be false. For the relata do entail that there are intentional psychological states, yet the relation is not instantiated by any such states. For example, if no one ever had had any intentional psychological state, 'being audited by the IRS' would not denote anything, but neither your being audited by the IRS, nor the relation between deducting your home office and being audited by the IRS, is instantiated by anybody's intentional psychological state.

There are endless further examples of intuitive causal connections that seem unproblematic and that are not instantiated by intentional psychological states, yet are not wholly independent of anybody's having epistemological or doxastic states either: Smith's failing French caused him to be ineligible to play NCAA Division I basketball; Milken's purveying of junk bonds weakened the US economy. None of these events — failing French, being ineligible to play NCAA Division I basketball, purveying junk bonds, or weakening the US economy — nor the relations between them, are instantiated in any intentional psychological state; yet, they could not occur if there were no intentional psychological states.<sup>26</sup> Therefore, on the conception of causation that underwrites CCP, the claim that there is a causal relation between, say, Smith's failing French and his ineligibility, is false. Since intuitively there is a causal connection, I do not think that the conception of causation that underwrites CCP is what we want anyway. So, let us reject that conception.

How, then, are we to understand causation? My suggestion is to take as our philosophical starting-point, not a metaphysical doctrine about the nature of causation or of reality, but a range of explanations that have been found worthy of acceptance. These include, pre-eminently but not exclusively, scientific explanations. They also include commonplace explanations that explain the phenomena that we encounter in everyday life — answers to questions like: 'Why is the traffic so snarled today?' 'Why is General Sherman's statue so shiny?' 'Why didn't you bring more beer?'<sup>27</sup> Construing explanations as answers to 'why' questions, with

<sup>26</sup> Eliminative materialists like Stich (1983) recognize that intentional action is as much at stake as belief and desire.

<sup>27</sup> For a defence of the explanatory worth of such explanations, see Baker (1987) and 'The Cognitive Status of Common Sense', in preparation.

perhaps some constraints on what can count as an adequate answer, my proposal is to begin with explanations that earn their keep, rather than with the metaphysics, which seems to me a freeloader that just interferes with real work.

Let me suggest that causes are the sorts of things that are cited in explanations of events. Let us understand 'c caused e' in this (loose and intuitive) way: (i) If c had not occurred, then, other things being equal, e would not have occurred, and (ii) given that c did occur, then other things being equal, e was inevitable (cf. Haugeland 1983). Whether one thing caused another, then, depends on what 'other things' get held constant; and in different explanatory contexts, it is appropriate to hold different 'other things' constant.

If we put aside the metaphysical picture and begin with the explanations that work, causation becomes an explanatory concept. This presents a sharp contrast to the metaphysical picture, which subordinates explanation to causation, where causation, in turn, is conceived as an 'objective relation' in nature. According to the metaphysical picture, causal power flows 'upward'—to put it metaphorically (how else can it be put?)—from subatomic particles through atoms and molecules through simple organisms through intermediate configurations up to persons with beliefs and other intentional states. There is no 'downward' causation. In the metaphysical picture, the font of all causality remains at the bottom.

If we reverse the priority of explanation and causation that is favoured by the metaphysician, the problem of mental causation just melts away. We began with the question: Does what we think ever affect what we do? Then, in light of the metaphysical picture, the question was recast as this: How can content-properties of internal events be causally relevant to producing behavioural events? With the reversal of priority of cause and explanation, the metaphysical version of the question just does not arise, and the original question has an easy answer. For example, when Jill returns to the bookstore to retrieve her keys, what she thinks is that she left her keys on the counter and that she wants them back. What she thinks affects what she does in virtue of the following explanatory fact: if she hadn't thought that she had left her keys, then, other things being equal, she wouldn't have returned to the bookstore; and given that she did think that she had left her keys, then, other things being equal, her returning was inevitable.

If we take our ontological cue from our successful explanatory

and predictive practices, then, admittedly, we end up with an ontological hodgepodge: statements concerning statutory laws, social roles, political, economic, and biological facts, as well as reasons, find their way into successful explanations and predictions. We make our reductions where we can, but on my proposal, we do not hold successful explanations hostage to ultimate assimilation into science. Unity is merely desirable, not inevitable.

My proposal, then, is to dismantle the problem of mental causation by rejecting the metaphysical background picture that generates it. If we accept paradigm cases of explanation in the sciences and in everyday life, and if we take the notion of explanation to be prior to that of causation, then the idea of a 'complete cause' in CCP hardly makes sense.

What we are left with is not dualism, nor materialism, nor any other comprehensive metaphysical doctrine. These doctrines do not earn their keep. Explanations do, in particular, explanations of what we do in terms of what we think. We could have almost none of the kinds of institutions and interactions that we have without such explanations. As for psychology, our relaxed attitude places no *a priori* constraints on the taxonomy of psychology. Indeed, different psychological theories, equally well-confirmed, may have different taxonomies, and thus systematize and explain different ranges of reality. Systematic explanatory success, in either science or everyday life, stands in need of no metaphysical underpinning.

## 5. Conclusion

Given standard metaphysical and methodological assumptions, not only has the problem of mental causation proved to be intractable but even worse: the same reasoning that leads to scepticism about mental causation also leads to scepticism about almost all supposed 'upper-level' causation, and hence to scepticism about explanations that mention 'upper-level' properties, including explanations offered by the special sciences and much of physics. Of course, pointing out such sceptical conclusions, even of this magnitude, is not a refutation of the metaphysical assumptions that generate them. But sceptical consequences may well be a motivation for taking a different philosophical tack.

Much of contemporary philosophy simply assumes the adequacy

of a metaphysical picture, and then takes the task of philosophers to be to determine what fits in and where it belongs in the picture. (We get all the news that fits.) Anything that resists such packaging is deemed unreal or illusory. But who said that reality had to be tidy? My proposal is to perform a methodological about-face. Instead of beginning with a full-blown metaphysical picture, we should begin with a range of good explanations, scientific and commonsensical. In the spirit of G. E. Moore, I think that our grounds for the claims that reasons sometimes explain behaviour are much stronger than any grounds for a metaphysical premiss that would lead to a contrary conclusion. And I am equally confident that our grounds for claims that macro-physical properties sometimes explain events and their properties are much, much stronger than any grounds of a metaphysical picture that would lead to a contrary conclusion.

Although my proposal has a strong pragmatic cast, it is by no means an anti-realist suggestion. I am not equating what is real with what is needed for explanations and predictions. The point is, rather, that we have no better access to reality than what is required for cognitive success, construed broadly enough to include what is cognitively required for achieving goals in both science and everyday life. Start with successful explanatory practices and let the metaphysics go. At least, we can avoid the insoluble problem of mental causation.