Rough draft chapter 19/2/04

Gabriel Segal

Comments invited

Content and Causation

(0) Introduction: the problem of mental causation

Allow me to recapitulate some territory that will be familiar to most readers. Here is how the problem of mental causation has typically been set up since shortly after the onset of non-reductive physicalism. It is now widely assumed that the realm of the physical is causally closed: every physical event has a complete physical cause, a cause that is sufficient for the event's occurrence. This apparently leaves us with a limited number of options concerning psychological causation, none of which appear hugely attractive. Either: (a) the psychological is epiphenomenal and can have no causal impact on the physical, or (b) the psychological is identical with the physical, or (c) thoughts and actions are all over-determined, each one having two distinct sufficient causes. Option (b) subdivides into two further options. Either (b1) the psychological reduces to the physical and every psychological property is identical with some physical property, or (b2) token psychological events are identical with or constituted from token physical events but psychological properties are not identical with physical properties. (b1) is widely held to be inconsistent with the multiple realisation of the psychological by the physical. And (b2) appears to bring us back to the original problematic, with the properties as the locus of tension. If one event causes another it does so in virtue some of its properties and not others. If I throw a stone at a window and the window breaks, it is because the stone was hard and heavy that it broke the window and not, say, because it was grey and millions of years old. The properties in virtue of which an event has a particular effect are typically called the 'causally efficacious properties of the cause with respect to the effect.' Suppose, then that token neural event causes an action. We can ask 'Does it do so in virtue of its physical properties or its psychological properties?' and we are back to choosing between options (a) and (c) or returning to (b1). And none of those options appeal.

That, as I say, is how the problematic is typically set up within the Davidson-Fodor package of token identity, type dualism (or type pluralism) and supervenience. I will call this, 'The Standard Package'. I think there is quite a lot that is seriously questionable about that way of looking at things. I want to accept the standard package and take a look at the problematic. My plan is as follows. First, I will say a few words about why I think that neither (a), epiphenomenalism, nor (c), overdetermination, should be rejected as quickly as they usually are. Indeed, these options tend to be dismissed so quickly that it is difficult to find articulated arguments against them. Kim (1998) is an exception and I will use his arguments as my focus. Kim's rejection of (a) and (c) lead him to favour a form of (b), reductionism. I will go on to criticise Kim's reductive approach. I will then turn to a proposal due to Segal and Sober (19xx) who defend a version of (c), overdetermination, and argue that the proposal does not work. I will then go on to argue for a version of (a), epiphenomenalism.

(1) The Standard Package and overdetermination

Let me begin with the standard package - in particular Fodor's version from Fodor (1974 'Special Sciences...') - and why it might it seem fitting to add overdetermination to it. Let us suppose that we have a (non-strict) psychological law to the effect that every event with psychological property F causes an event with psychological property G; "Ceteris paribus, every time someone has a strong desire to jump up and down, they jump up and down" for example. F has a wide variety of possible physical supervenience bases, $m(F)_1$, ..., $m(F)_n$. Every time the law is instantiated, the occurrence of the F event is realised by the occurrence of some m(F) event. The m(F) event causes, by physically explicable, unmysterious mechanisms, the occurrence of some m(G) event, where m(G) is a supervenience base for G.

Let us for the time being just accept that metaphysical picture. Then it is tempting to say that just as paradigmatic psychological properties (e.g. being a desire to jump up and down) supervene on physical properties, so also psychological laws supervene on physical laws and psychological causal relations supervene on physical causal relations. And, if we are tempted to say that, then we might well be tempted by overdetermination. For we do not conclude from the fact that psychological properties supervene on physical ones that psychological properties don't exist. We do not conclude from the fact that psychological laws supervene on physical laws, that there are no psychological laws. So, to be consistent, we should adopt a similar attitude to psychological causation: we should not conclude from the fact that psychological causation supervenes on physical causation that it does not so much as occur. We have independent reason to believe that the m(F)s are causally efficacious in respect of causing G events. But that should not prevent us from allowing that F itself is also causally efficacious in respect of the very same effects. We should not think of F and the m(F)s as competing for efficacy. Rather, we should think of F's efficacy as inherited from its m(F)s.

So, on this approach, both m(F) and F are causally efficacious in the matter of causing Gs. That means that they are both sufficient for G events, under the circumstances. And it is that kind of overdetermination that a lot of people find very hard to believe in. But why is it so hard to believe? Of course, if it were merely a coincidence that every time an F event caused a G event, some m(F) event or other just happened to crop up, for no apparent reason, and cause that same G event (as it were, a second

time over), then that would indeed be incredible. That would be analogous to a world in which every time somebody was killed by a bullet they were also, at the same time, killed by some other unrelated means. But that is not how it goes with the standard package. The occurrence of F events and m(F) events are, of course, not unrelated. Rather, the co-occurrence of Fs and m(F)s is guaranteed by the supervenience component of the package.

On this view, overdetermination is not rare, but ubiquitous. It holds in all cases that fit the metaphysics of The Standard Package. If I throw a stone at a window and the window breaks, then we can explain what happened at more than one level. For example, we can explain it in relatively macroscopic terms appealing to things like the stone's hardness and weight. And we can explain it at the microscopic level by appealing to things like the arrangement of the stone's component atoms and the strength of bonds among them. Both explanations cite causally efficacious properties of the stone: e.g. its being hard and its being so-composed of atoms. The explanations don't compete. So why suppose that the properties compete for the title of 'causally efficacious'? Don't prejudge the question of whether overdetermination is ubiquitous, but look and see whether it is.

I think it is fruitful (following Yablo 19xx) to think about the issue in relation to determinates and determinables. Suppose we place a crimson cube in front of a mirror. This causes an image to appear to in the mirror. The image has many properties. It is crimson and, let's say, square. But it is also red and rectangular.

Now what properties of the object caused there to be a red, rectangular image?

One obvious answer is: redness and rectangularity. But another perfectly good answer is: crimsonness and squareness. These properties of the object cause there to be a crimson square image. And, since it follows logically that the image is also red and rectangular, they cause there to be a red, rectangular image. There is nothing very mysterious about this. Both crimsonness and redness, for example, have the power to cause red images, and the reason they do their causing at the same time and in relation to the same effects is because being crimson is a way of being red. Moreover, each time a red thing causes a red image, the redness is realised by some determinate shade that causes a correspondingly shaded image. But there is nothing particularly mysterious about that either. There are just many ways of being red and every red thing has got to be red in some way or other.

(2) Kim and Reductionism

Kim regards the overdetermination approach to mental causation as 'a non-starter' and he enters an objection to it. Since it somewhat cryptic, I quote the entire passage (Kim p. 45):

> consider a world in which the physical cause does not occur and which in other respects is as much like our world as possible. The overdetermination approach says that in such a world the mental cause causes a physical event - namely that the principle of causal closure of the physical domain no longer holds. I do not think that we can accept this consequence:

that a minimal counterfactual supposition like that can lead to a major change in the world.

But what is this 'minimal counterfactual supposition'? Take a particular case in which an F event causes a G event. Suppose that in this case F's base is $m(F)_j$. And consider the closest possible world, W, in which the $m(F)_j$ event does not occur. What does the overdetermination approach imply about W? Kim says it implies that 'the mental cause causes a physical event'. We need to be a little bit careful here. If the physical cause is a token event, then the most natural view would be that the event that is the mental cause in the actual world does not even occur in W, since the physical event that it is identical with, or constituted from, does not occur in W. However, it seems reasonable to suppose that some other F event occurs in W, perhaps one very similar to the F event that occurred in actuality. And that other F event would indeed cause a G event in W. But, of course, that doesn't entail any violation of the causal closure of the physical. F has various possible physical base m(F)s, and some m(F) event or other (perhaps even some $m(F)_j$ distinct from the actual one) would have occurred to constitute or be identical with the F event.

In fact, related counterfactuals help to support the overdetermination approach rather than threatening it. If an F event had occurred and an $m(F)_j$ event not occurred, then the F event would still have caused a G event. That surely lends some prima facie support to the idea that F is causally efficacious with respect to Gs, even if we have independent reason to hold that $m(F)_j$ is causally efficacious in respect of Gs as well.

So let us not dismiss the overdetermination approach too quickly. It is, at least, a starter. What, then, about epiphenomenalism?

Epiphenomenalism about psychological properties has had some champions. But still, as far as I can tell, there is a bit of a tendency for people to dismiss it too quickly. Some think that the causal efficacy of psychological properties is given in our experience. Surely I know why I jumped up and down: it's because I wanted to. It is true that some of the time, it does seem to us that we know why we do something. I might contemplate whether to jump and down, be fully aware of my desire to do so and then, in the light of this desire, go ahead and jump. In such a case, it might well seem to me that my desire to jump up and down causes me to jump up and down. But, in my case at least, I don't feel that I have a particular sense of which properties of that desire caused the action. Nothing in my experience contradicts the idea that it was its physical properties, and those alone, that did the work.

Kim cites a case that might seem more compelling: pain. Suppose that I suddenly feel a burning pain, as I accidentally lean on the hob. I instantly pull my hand away. Surely, someone might say, it is because your hand hurt that you moved it. But even in that case, I don't think my experience presents me with a view about what properties are doing the causal work. I feel the pain and I move my hand, but I do not experience the painfulness of the pain causing the hand to move. It is Hume all the way.

In fact, both experience and science offer some support to epiphenomenalism about conscious states. For sometimes we feel the pain after we have performed the action that we might otherwise think was motivated by it. And science tells us that sometimes the neural events that initiate the causal processes of

our actions occur before we know what we are trying to do.

Many people also seem motivated by the idea that it would be unfortunate - even terrible - if epiphenomalism were true and that some of our treasured views about ourselves and our place in nature would be under threat. Kim gives voice to three concerns of that sort. If we keep clearly in mind that it is the causal potency of psychological properties that is at issue, I think that we can see that the concerns are groundless.

First (p. 31): 'the possibility of human agency evidently requires that our psychological states ... have causal effects in the physical world'. A little clarification is in order here. It does seem right that my agency requires that I do things because of what I want, what I believe and so on. And it seems right to take that as a causal 'because'. So, for example, if I jump up and down just because I want to, it is my desire to jump up and down that causes me to jump up and down. But that is token-event causation. What is not at all evident is that our conception of human agency requires further that the psychological properties of my desire be causally efficacious in moving my body.

Second (p 31): the possibility of human knowledge presupposes the reality of psychological causation. For example, Kim says, reasoning requires the causation of beliefs by beliefs and memory requires the causation of beliefs by perceptions. I don't think that is plausible at all. The acquisition of knowledge is compatible with epiphenomenalism even about token mental events. What knowledge requires is that the processes by which beliefs are formed be reliable, not that they be psychological. That is part of the point of automated computation. In a computer, physical processes ensure the production of representations that stand in appropriate rational relations to one another, given their interpretations. If we are ensembles of computers, as cognitive science supposes we are, then we could be good knowledge-acquisition devices whether or not our representations' psychological properties have causalpotency. Suppose, for example, that one of a certain creature's belief-forming mechanisms is a physically implemented truth-preserving inference machine that only creates a new belief if it is a logical consequence of old ones. If other conditions are right, then this creature would be in a good position to acquire new pieces of knowledge via the operation of the mechanism, epiphenomenalism or no.

Third (still on p. 31): 'the possibility of psychology as a theoretical science capable of generating law-based explanations of human behaviour depends on the reality of psychological causation: psychological phenomena must be capable of functioning as indispensable links in causal chains leading to physical behaviour'. Again, this might be right if we are talking of psychological events and processes. Many existing branches of psychology are committed to causal psychological laws. The laws relate types of psychological phenomena to their effects. But these sciences are not committed to telling a complete story about which properties of the causes are responsible for bringing the effects about. This is clear, for example, in classical cognitive science. In theorising about a given cognitive system, a cognitive scientist aims to tell us three stories about which representations will cause the production of which other representations: semantic, syntactic and physical (e.g. neural). Prima facie, given the way most of us tend to think about causation, it looks as though cognitive science does aim to tell us something about which properties of representations are causally responsible for their characteristic effects: the physical ones. (I am not convinced that even that much is right. But let that pass.) But what cognitive science emphatically

does not tell us is whether the syntactic and semantic properties are causally efficacious in respect of the same effects. That is why the matter is still debated by philosophers. Existing psychology leaves open questions about the efficacy of psychological properties. And future psychology will be able to accommodate the causal role of psychological properties, whatever that role turns out to be.

Kim's rejection of overdetermination and his epiphobia lead him to try and formulate a plausible version of reductionism. Like most of us, Kim accepts multiple realisation. The proposal he outlines (which is similar to David Lewis's brand of funcitonalism, which might better be termed 'physico-functionalism', Lewis (19xx)) is as follows. (Kim does not fully endorse the proposal but rather expresses hope that it might turn out to be right). On the view Kim proposes, psychological descriptions are equivalent to second-order functional descriptions. So, for example, to satisfy some mental predicate 'M' is to have some first-order property, P_n, such that P_n has some particular characteristic causal role. P_n will be a physical property, in effect a realiser (an m(M)) for the mental property M. Kim assumes that M has multiple physical realizers in different 'species and structures and can have different realizers in different possible worlds' (110). He continues (110):

> The reduction consists in identifying M with its realizer P_i relative to the species or structure under consideration (also relative to the reference world). Thus M is P_1 in species 1, P_2 in species 2, and so on. Given that each instance of M has exactly the causal powers of its realiser on that occasion ... all the causal explanatory work done by an instance of that occurs in virtue of the instantiation of realizer P_1 is done by P_1 ...

This offers an illusion of saving the efficacy of mental properties. Here is the illusion: we have a mental property M, in a species or structure k, at a possible world W_j, and that property is identical with some P_k and so, unproblematically, has P_k's causal powers. A couple of clarifications should bring out why it is an illusion. First, 'species or structures' are those items that have just one way of realising M at any given world. Maybe talk of species is here appropriate for some types of mental property, like, say, the sort of pain caused by burning. (Recall however, that Kim is not optimistic about the possibility of the story applying to phenomenal properties). And it might be appropriate for certain representational properties as well, such as representations in hard-wired bits of computational modules. For such things, maybe, M has a single realizer across a whole species. But talk of species is not appropriate for lots of interesting psychological states. Consider propositional attitudes, for example. It is very unlikely that a given propositional attitude, such as the belief that Barcelona is beautiful, will have the same physical realisations in different people, or, perhaps, in the same individual at different times. So, at least as far as propositional attitudes are concerned, 'structures' are likely to be individual cognizers, or cognizers-during-periods, at particular worlds.

The second point of clarification is that what get identified with the P_ns are not mental properties like believing that Barcelona is beautiful. They are properties that might be described along the lines of: M-in-

structure-K-in-W_j. And of course M-in-structure-K-in-W_j, M-in-structure-K-in-W_i and M-in-structure-J-in-W_j etc. are all different properties. But then such properties are not psychological. If you and I and Genoveva all believe that Barcelona is beautiful, then we share a psychological property. There are no further psychological properties corresponding to my-actual-belief-that-Barcelona-is-beautiful and your-actual-belief-that-Barcelona-is-beautiful etc..

In fact, psychological properties suffer a worse fate than mere impotence, on Kim's view. Psychological descriptions are to be reconstructed as functional descriptions, so, at first pass, we would construe psychological properties as functional properties. But Kim doesn't really believe in functional properties, since he

makes a case for "eschewing the talk of functional *properties* in favor of functional *concepts* and *expressions*." According to Kim, functional concepts and expressions are perfectly good and useful. But they don't pick out real properties: or at least they don't pick out 'robust' properties of the sort that feature in proper scientific generalisations.

I think that the fate of Kim's package reveals the dangers of believing in multiple realization while rejecting both overdetermination and epiphenomenalism.

It very much looks as though multiple realization is just of those things we will have to accept. So it seems sensible to see if we can make a reasonable case for one of the other options. I will now outline Segal and Sober's defence of an overdetermination approach, and argue that it fails. I will then offer an alternative that I shall present as a version of epiphenomenalism.

(3) Segal and Sober

Segal and Sober offer a sufficient condition for the causal efficacy of a macro-property, which they label (P5), and argue that psychological properties meet that condition. I will state (P5), which is a touch complicated, and say a little about the motivation behind it.

> (P5) If (i) it is a (possibly nonstrict) law that every F event causes a G event and (ii) in each case in which an F event causes a G event there exist micro-properties m(F), m(F)' and m(G) such that the cause's being F mereologically supervenes on its being m(F) and the effect's being G supervenes on its being m(G) and possession of m(F) includes possession of m(F)' and the cause's being m(F)' causes the effect's being m(G), then F is efficacious in the production of Gs.

(P5) is an attempt to explain macro-causation in terms of micro-causation. The idea is that the combination of nomological and supervenience requirements should be strong enough to rule out all counterexamples. It is well known that laws aren't enough: for example, successful matings of blue-eyed

individuals cause births of blue-eyed children. But the property of involving blue-eyed individuals is not causally efficacious in respect of the mating's production of blue-eyed children. Rather, it's a property relating to the parents' genes that causes both the parents and the children to have blue eyes. Segal and Sober's thought was that correlations of this sort could be ruled out by requiring a sufficiently tight relation between the macro-property of the cause and the micro-properties that we assume to be doing causal work in bringing about the effect. Mereological supervenience was thought to do the trick. Roughly speaking, mereological supervenience is the converse of the 'makes it the case that' relation that we talk about when we explain why an object has certain macro-properties by citing properties of and relations among its components (at some level of description): e.g. the way the diamond's crystals are bound together makes it the case that the diamond is hard.

In fact, even by Segal and Sober's own account, (P5) is too weak as stated. They require further that F be a 'substantial' property, meaning a property that pulls its weight in good scientific generalisations (a 'robust' property in Kim's sense). This rules out properties like being-red-or-weighing-a-hundred-pounds. They admit that they don't have a particularly clear or informative account of what makes for a substantial property.

(P5) caters specifically for cases in which the supervenience base of F is complex and includes distinguishable micro-properties. This is to allow for cases in which some but not all of the components of the base are involved in the causal transaction at issue. For example: smoking causes cancer. In fact, the activity of smoking involves all sorts of things that do not cause cancer. Only some of what is involved in smoking - the inhalation of material that has certain specific carcinogenic properties - does the work of causing cancer.

Segal and Sober go on to argue that mental causation fits that model: psychological properties supervene on complex neural (and perhaps other) properties, some of which are involved in the causal transactions of interest and some of which are not.

Segal and Sober note that their view entails that many second-order functional properties are causally efficacious with the respect to the effects relative to which they are defined. This leads them into a dispute with Ned Block. Block allows that second-order functional properties might be efficacious in respect of certain effects in which intelligent beings recognize them. But he denies that they are efficacious in respect the specific effects that define them. He discussed a bullfighter's cape. Since the cape is red, it is provocative to bulls (let us suppose, he says, for the sake argument, even though it is not in fact true). The bullfighter uses the cape to provoke the bull. Block claims that it is the redness of the cape that caused the bull to be provoked, and not its provocativeness.

Segal and Sober voice the suspicion that Block has confused properties with ways of referring to properties. They draw on the thought that we should not think that a property is inefficacious with respect to a certain effect merely because we pick it out by adverting to that very effect. They discuss the example of solubility, saying this:

we might construe solubility as a second order property, defined thus:

(i) x is soluble iff x possesses some property F such that x's being F causes x to dissolve when immersed.

But solubility may be defined in other ways. Suppose that there is a unique substantial property that causes objects to dissolve. This property may be defined as follows:

(ii) solubility is the property that causes objects to dissolve when immersed.

So (i) and (ii) provide alternative definitions of the same property. It follows that second order properties may have effects on stupid objects. Of course, there may be no substantial property defined by (ii). But this is what Block would need to show to establish his claim.

But it is Segal and Sober and not Block who display confusion in matters relating to properties and ways of referring to properties. I take it that a unique substantial property that causes objects to dissolve when immersed, would be a micro-property.

Definition (ii), if it defines anything real (which I doubt), defines that micro-property. But the existence of such a micro-property is metaphysically contingent. Surely, even if solubility has only one nomologically possible realizer, it has many metaphysically possible ones. But then some metaphysically possible objects are soluble in sense (i) but not in sense (ii), so (i) and (ii) do not define the same property. And that makes trouble for (P5). For even if a certain range of functionally specified properties, namely those exemplified by solubility in sense (ii) (if there is such a property) can be efficacious in the relevant respects, (P5) incorrectly licenses the efficacy of others.

I think Block was right. Functional - that is, genuinely dispositional —properties, whether second-order or not, are not efficacious with respect to their defining effects. Let me reformulate (i) slightly as (i’):

(i’) x is soluble iff (a) x would dissolve if immersed and (b) x would be caused to do so by some property F.

Now, it is part of the notion of a causally efficacious property that it enters into the causal explanation why its possessors bring about their effects. A property that is efficacious in respect of causing an object to dissolve in water has to enter as a causal factor in the explanation of why the object would dissolve, if immersed in water. Being soluble in sense (i’) can’t do that. An object’s being soluble in that sense just *is* (in part) its being such as to dissolve when immersed.

Segal and Sober repeat Davidson’s remark to the effect that statements like ‘the cause of A caused A’ are uninformative but, of course, not false. They do this to emphasise the point that just because our description of a cause relates it logically to an effect does not mean that it is not the cause of that effect.

But properties differ from ordinary individual objects and events. Properties can be essentially dispositional, ordinary individuals and events can't. The cause of A is not metaphysically related to A: it might have occurred without A having occurred. But an object can't be soluble in sense (i') and yet not dissolve if immersed. Moreover, the cause of A has other properties than merely being the cause of A, and its possession of some of those explain why it caused A. And while one can describe solubility without talking of objects dissolving, one cannot use such descriptions in an explanation of why solubility causes objects to dissolve. There is no such explanation.

It is not the metaphysical connection between property and effect that is the problem. It is because an object's solubility consists in its being such that it would dissolve if placed in water, that solubility cannot be a causal factor in making objects dissolve in water. I will say more about dispositions and causal efficacy in the next section. For now, note that (P5) entails that lots of dispositional properties are causally efficacious in respect of their defining effects; for example solubility, fragility and elasticity, all considered on the model of (i'). So (P5) won't do.

If one wanted follow up on Segal and Sober's start, one could perhaps deal with the problem by adding a further constraint: where F is a substantial and not essentially dispositional property, (P5). This is not the path I will take, however, partly for reasons that will emerge in the next sections and partly because there is another problem with (P5).

Consider the particular sort of red glow that red-hot pokers get. It is likely that it is a non-strict law that objects with that very specific sort of red glow cause wax to melt. The red glow supervenes on the agitation of the objects' molecules, and it is this that causes the agitation of the wax's molecules which, is the supervenience base of the wax's melting. But the red glow is not causally efficacious in respect of melting the wax.

One might doubt that it is a law that objects of that colour melt wax. Maybe a really good special-effects person could create objects that have that colour at room temperature. I don't know. But we can still imagine a possible world in which there are mechanisms that systematically prevent the creation of any such objects, so that there, it is a ceteris paribus law that objects of the requisite colour melt wax. But still is not because objects have that colour that they melt wax.

Before continuing with the issue of mental causation in particular (section 5, below) I need to say more about the metaphysics of properties, dispositions and causation.

(4) Properties, Dispositions and Causation

Let me begin by describing an extreme view. It is a view that Sidney Shoemaker briefly canvassed, but quickly abandoned. It is a view that most people find incredible. But it is a view that has a great deal to be said for it and that is very difficult to refute. The view concerns those properties that feature in causal explanations. It thus has broad but not universal application, failing to apply to properties of abstract objects, such as being a prime number.

The rough idea is that properties are just clusters of dispositions. So, for example, what it is for an object to be spherical is for it to be such that if placed on a slope, it would roll down; if placed in certain relations to a mirror it would cast a round image; if placed on plasticine it would create a hemispherical hollow; and so on and so forth.

Shoemaker offers a refined version which identifies properties not with dispositions but with what he calls 'conditional powers'. Dispositions relate objects to effects in circumstances: e.g. to be soluble is to be such that you dissolve, if placed in water. Conditional powers relativise dispositions (powers) to other properties the object might possess. So, for example, a spherical object has the power to produce a round image, if placed in certain relations to a mirror, if it has a reflective surface. And it has the power to produce a hemispherical hollow if placed on plasticine, if it is hard and heavy.

Let us call the view that properties are clusters of conditional powers 'the extreme causal theory'. According to the extreme causal theory, properties are individuated by the causal powers of their possessors: the nature of a property is exhausted by the effects its possessors would bring about, given other properties they might have. Here are two considerations which lend support to the theory. The first is from Shoemaker (2003 215).

Suppose first, for reductio, that properties X and Y are not individuated by the powers of their possessors. Then you could get two objects that were utterly indistinguishable by any possible test but yet differed in some of their properties. But then we could never know anything about these different properties. Nor could we say anything true and interesting about what distinguishes them. So it is sensible to conclude that F=G iff possessors of F and possessors of G have the same conditional powers, hence that properties are individuated by the powers of their possessors.

Here is a second, more complex consideration. Suppose I throw a stone at a window and the window breaks. Why did the window break? Material science tells us about that. The stone is made of stiff material (that is to say the material has a high 'Young's modulus', or resistance to deformation). The stone is massive and spherical. And it is stronger than the glass (that is to say it has a higher 'work of fracture', or impact strength; roughly, resistance to breaking under impact). The stone is travelling fast and a small area of its surface comes into contact with the glass. The glass is thin, relatively weak etc., and it breaks.

Now it is obvious (on inspection, I claim) that a number of the properties featured in the explanation are purely dispositional, hence individuated by the powers of their possessors. Stiffness and impact strength are paradigm cases: for the stone to be thus stiff and thus strong *just is* for it to be such that, given its other properties, it would break such a thing as the window, under suitable impact. So, suppose you have objects somewhat similar to the stone and the window and you throw the former at the latter and the stone breaks, rather than the window. Investigation reveals that their shapes, Young's moduli and every property other than their works of fracture are just the same as the originals. It follows logically that the work of fracture differs in at least one case. Either the second window had higher work of fracture than the first window, or the second stone had lower work of fracture than the first stone, or both.

Now it is tempting to suppose that not all of the properties featuring in the explanation are similarly related to the causal powers of their possessors: the stone's shape, for example. It does not seem that being spherical is in any essential way related to the capacity to break windows.

But further reflection shows that the tempting supposition is mistaken. To give a proper account of what having a certain degree D of strength is, we have to talk about the shapes of strong things. But if we are to treat D as a conditional power, thus properly accounted for, then shape is metaphysically implicated in the causal nexus. Strength can only be essentially related to the effects of its possessors if shape is too. There is no metaphysically possible world W in which the stone has all its actual properties but fails to break the window because sphericality is associated with different conditional powers in W. It is, therefore, metaphysically necessary that spherical things that share the other properties of the stone, break things such as the window. If there is a world W1, in which the stone doesn't break the window, but is still spherical, then it follows that in W1, either the stone or the window differs from how it is in W in respect of at least one property: either the stone is less stiff or strong or the window is stiffer or stronger or whatever. It is by the causal relations holding among their possessors that such properties, the dispositional ones like stiffness and strength, are defined.

So, it seems, shape is just as intimately connected to the causal powers of its possessors as is stiffness or strength. On the face of it, it looks as though strength is essentially connected to the tendency to break while sphericality is not. But further thought reveals that both properties stand in metaphysically necessary relations to the effects of their possessors. The difference is merely epistemic.

The connection between, say, fragility and breakages is relatively obvious to us. It is conceptually necessary that fragile things break, when hit hard. Or, to put it better: it is conceptually necessary that fragile things break when hit hard enough by other things of the right sort. Material science spells out and quantifies tautologies of that ilk, replacing the crude common-sense notions of fragility and the like with technical, quantitative ones that fit the contours of the real world.

The connection between sphericality and breakages is not obvious to us at first glance. There is nothing in the mere concept of sphericality, considered by itself, that immediately relates it to breakages. However, I do think that the material scientific explanation of what happens when a spherical stone breaks a window articulates a conceptual necessity. Once you understand the concepts involved, you will see that a thing with the properties of the stone would break a thing such as the window, under the relevant circumstances. The explanation includes no appeal to conceptually or metaphysically contingent laws of nature. It just follows from the correct account of the properties of and relations between the two things, that the stone would break the window.

So, the extreme causal theory has two virtues. First, it avoids the postulation of utterly mysterious properties the possession of which makes not the slightest difference to anything. Second, it accounts for the fact that paradigmatic causal explanations articulate metaphysically necessary relations among properties.

Here are three objections to the extreme causal theory.

The first is from Lewis (1986). Lewis adopts a 'principle of recombination': (88) 'roughly speaking the principle is that anything can co-exist with anything else'. The principle expresses a version of the Humean idea that there are no necessary connections between distinct existences. And it naturally leads to the conclusion that laws of nature are not metaphysically necessary, for events that are connected by natural law are distinct existences, and hence are mixed and matched in all different ways across possible worlds. Events of stiff, strong stones hitting windows are distinct from events of windows not breaking. Hence there is a possible world where events of the former kind are followed by events of the latter kind. So the principle of recombination entails that properties featuring in natural laws are not metaphysically connected to the powers of their possessors after all (Lewis 1986 163). Lewis anticipates the possibility that his opponent will simply reject the principle of recombination and responds that he would thereby be leaving the frying pan for the fire. He omits to say which fire.

But there is no clash between the principle of recombination and the extreme causal theory of properties. There are surely lots of worlds in which people like me throw things like the stone at things like the window and in which the window does not break. The extreme theorist can allow that there are as many of these worlds as Lewis wants. That is not the issue. The issue is how we should describe such worlds. Lewis wishes to describe them as worlds in which stiff, strong stones fail to break windows. The extreme theorist wishes to describe them as worlds in which the stones or the windows or both have properties different from their actual ones: the glass, for example, might be stronger or the stone less strong. Hence Lewis's objection misses the mark.

The second objection is due to Swinburn and Armstrong. I find the objection a little hard to formulate, but the idea seems to be that the account is viciously regressive or that it leaves properties ungrounded, in some sense (Armstrong 31):

> Every causal transaction, according to Shoemaker, is a matter of things with certain potentialities bringing it about that these or other things have further potentialities, because properties are analyzed as nothing but potentialities. In Scholastic language, we never get beyond potency to act. Act, so far as it occurs, is just a shifting around of potencies … 'Where's the bloody horse?' as the poet Roy Campbell might have asked.

But this is a statement of incredulity rather than an objection. Suppose that I make a snowball. Since I am strong, I can exert pressure on the snow and cause the snowball to be hard. It is pretheoretically plausible that my strength and the snowball's hardness are both dispositional properties. So this is a case of one potentiality being involved in the creation of another potentiality. And there is nothing mystifying about it at all. We understand what is going on. It is a consequence of the extreme theory that causation is always like that. And why shouldn't it be so? If it can happen some of the time, it can happen all of the time.

The third objection can also be found in Armstrong (19xx). It ought to be possible for there to be a universe with a totally symmetrical nomic and causal network, in which each property has a mirror image. According to Armstrong, there is no way in which the extreme theorist can distinguish two symmetrically

opposite properties in such a universe. But that is not true. Consider this example from Denis Robinson. Imagine properties F and G: particles acquire these properties in well defined circumstances, with a 50% chance of getting either. If two particles both have F or both have G they repel each other: if they have one of each, they attract each other.

Suppose, for the sake of argument, that F and G are purely dispositional. That does not make them identical! The disposition to attract a G particle is quite different from the disposition to attract an F particle, even F and G are mirror images. Of course, we need to suppose that attraction and repulsion are images too (as are large and small, further and nearer etc.). That still doesn't make them the same. Suppose we lived in such a world. We might point to two particles that were attracting one another and define predicates 'F' and 'attracting' as what those particles are doing now.

Armstrong, above, attributes the extreme view to Shoemaker. But in fact Shoemaker abandoned it in the very article in which it first made its appearance. I can discern two reasons why Shoemaker didn't stick with it. The first is that he wants properties to explain causal powers. An object's powers, he thinks, are 'grounded in' its intrinsic properties (2003 213). If that is right, then obviously properties can't be identical with powers. So Shoemaker's ultimate view has properties endowing their possessors with causal powers, rather than just being those causal powers. So, for example, a window's fragility is supposed to explain its tendency to break, rather than to be that tendency.

The second reason is due to Richard Boyd. It goes as follows (Shoemaker 2003 232-4):

Imagine a world in which the basic elements include four substances, A, B, C and D. Suppose that X is a compound of A and B and Y is a compound of C and D. We can suppose ... that the property of being made of X and the property of being made of Y share all of their causal potentialities.

It would follow from the extreme view that being made of X and being made of Y would be the same property. But Shoemaker finds that counterintuitive, since X and Y would be different substances. So, according to Shoemaker's ultimate theory, properties are individuated not just by their associated powers, but also by the circumstances that cause them to be instantiated. The ultimate theory is that property F=property G iff F and G endow the same conditional powers on their possessors and whatever circumstances suffice to cause and instantiation of F suffice to cause an instantiation of G and vice versa.

Notice that Shoemaker's ultimate theory has the two virtues by which I advertised the extreme theory. It avoids positing properties that make no difference to anything. And it explains the metaphysically necessary character of paradigmatic causal explanation. Nevertheless, I think the extreme theory is preferable to the ultimate theory. The retreat from the former to the latter is unwarranted on either count. Let us consider Boyd's example first.

There is no compelling reason to suppose that X and Y are distinct substances. It is at least as natural to suppose that there is just the one substance that can be created in either of two ways. Suppose that we lived in Boyd's world. Would we even bother to have two words for X and Y? That is not likely, since it

would be a major task to keep track of which samples had which origins. Moreover, we are trying to provide an account of properties that feature in causal explanation. From that perspective, there is no reason to distinguish being made of X from being made of Y.

One can of course distinguish X from Y if one wants to. But the distinguishing properties are the historical ones: having been created from A and B versus having been created from C and D. Those properties are perfectly distinguishable on the extreme causal theory, as long as A, B, C and D are themselves distinct in causal powers. And it is safe to assume that they are, otherwise Boyd's example would make very little sense. There is thus no need to think of X and Y as different substances. They are batches of the same substance that differ only in the manner of their creation.

Shoemaker's other reason for abandoning the extreme theory raises subtle issues. Let us consider fragility. Shoemaker sees fragility as an intrinsic property that grounds a disposition. I see it as the disposition. According to the Oxford English Dictionary, 'fragile' means: *easily broken*. My understanding of English leads me to accept that definition just as it is. The dictionary does not say that 'fragile' means: *the property in virtue of which easily broken things are easily broken*. And there is no good reason so to interpret what it does say along those lines.

Further, as I have been insisting, the scientific successors of folk notions like fragility or hardness are, likewise, purely dispositional. Terms like 'Young's modulus' and 'work of fracture' specify dispositions that can be precisely quantified. And material science explains how objects come to possess these dispositions. The explanations are micro-structural, citing properties of and relations among objects' constituents. Thus the property that explains why the window is easily broken is not its fragility (or low impact strength), but rather its being composed of such and such matter in such and such an arrangement. That the object is easily broken under impact is a purely dispositional matter. And it is thus easily broken because it has that particular micro-structure. There is no third property of fragility that somehow comes between those two, supervenient on the latter and explaining the former. Occam's razor is on my side. Fragility is the disposition, neither more nor less.

We do, of course, say things like: 'the window broke because it was fragile'.

This locution may perhaps mislead some people into thinking that fragility is a cause of the window's breaking. But the relevant 'because' means: *by reason of*. And reasons can be metaphysically, conceptually and logically intimate with what they explain in a way that causes cannot (in the *effect* sense of 'cause'). For example, consider a right-angle triangle, with sides A and B of 5 cms and 4 cms respectively. The third side, C, is 3 cms, because A^2-B^2=C^2 and 5^2-4^2=3^2.

If the extreme theory is right, then the material science explanation of the window breaking is not too different from the geometric explanation of why C is 3 cms. And there is precisely nothing wrong with that. The scientific explanations make transparent to us how various combinations of properties interrelate in the jigsaw puzzles of causal interactions. They have other virtues as well.

One reason why we might want to know, for example, that stiffness and strength are properties of the stone that are relevant to whether it would break a window is that we can identify its possession of those properties without having to see whether it breaks anything. You can feel them with your hand.

Another reason why we might be interested is that while it is conceptually necessary that something as stiff and strong etc. as the stone would break a weaker thing like the window, it is not conceptually necessary that only things with those properties can break windows. Under the right conditions, you could break a window with a large meringue or a mound of jelly. There are many different combinations of properties that can combine to make something able to break a window. So the macroscopic explanation of the event serves to rule out many other possible macroscopic explanations.

I have a suspicion that we start out, intuitively, with somewhat confused notions of certain dispositional properties and their role in explanation. We have terms like ‘hard’ that don’t map very well onto the scientifically interesting properties. ‘Hardness’ seems to cover the conjunction of impact strength and stiffness, though not tensile strength (resistance to being pulled to pieces). We can, to an extent, identify the presence of such properties in bodies by touch. And we think of them as the properties that produce the feelings by which identify them. That makes it appear as though they are contingently related to effects involving interactions among inanimate things. It appears, and perhaps is, contingent that an object that feels hard, in the manner of the stone, would break such a thing as a window. And that leads us to think of dispositional properties as playing a role in causal explanation that they cannot play. We think of the stone’s hardness as causally responsible for its breaking the window, while actually it is not.

The extreme theory has a lot going for it. But if it is right, then we should probably give up on the notion of a causally efficacious property. That notion appears to arise from misreading the ‘because’ in sentences like ‘the stone broke the window because it was stiff and strong’ as an *effect* ‘because’ rather than a *reason* ‘because’ and we think of the properties mentioned as causes of the effects they explain.

There is, however, a slightly less extreme view that does not put so much pressure on the notion of casually efficacious properties. Let us return to the contrast between sphericality (not obviously dispositional) on the one hand and, say, impact strength (more obviously dispositional), on the other. Suppose that the extreme view is right in claiming that both kinds of property have metaphysically necessary connections to the effects of their possessors. And suppose that these are mirrored by conceptually necessary connections articulated in canonical explanations of the properties’ possessors bringing about their effects. It does not quite follow that all the properties are individuated in terms of all of the conditional powers they are associated with. There remains room for the view that the stone’s strength is more intimately connected to its capacity to break windows than is its sphericality. For the stone to be as strong as it is, is (in part), for it to be such that, given its other properties, it would break such a thing as the window, if thrown at it. But maybe it is not true that for the stone to be spherical is (in part), for it to be such that, given its other properties, it would break such a thing as the window, if thrown at it. Its being spherical might entail that it would break a thing such as the window. But that does not mean that its tendency to break a thing such as the window is part of what makes it the case that it is spherical.

And that seems right. The stone's being spherical consists in its being such that every point on its surface is equidistant from its centre. And if that is right, then the stone's sphericality can be causally efficacious in respect of the window breaking, even if metaphysically and, in canonical explanations, conceptually, connected to it.

If the less extreme causal view of properties is right, then the "because" in "The stone broke the window because it was stiff and strong and spherical and ..." is a *reason* because, not an *effect* because. Nevertheless, some of the properties mentioned, such as sphericality, are also causally efficacious in respect of bringing about the effect.

The issue of the causalefficacy of mental properties now becomes the issue of whether they are like dispositional, like fragility or not sphericality. To this I now turn.

(5) The Standard Package and the Amazing Coincidence

Let us have another look at the standard package and its account of mental causation. It goes as follows. Suppose it's a law that F events cause G events, where F and G are psychological properties. F has a variety of physical supervenience bases, $m(F)_1$, ..., $m(F)_n$. Every time the law is instantiated, the occurrence of the F event is realised by the occurrence of some m(F) event. The m(F) event causes the occurrence of some m(G) event, where m(G) is a supervenience base for G. But if that is right, then it shrieks for explanation. Consider the m(F)s. They all share two properties: (a) they are bases for F and, further, (b) they all tend to cause the occurrence of a base for G.

That is a remarkable correlation and there must be some explanation for it. There must be a connection between being a base for F and tending to cause bases for G.

Suppose that you have an account of representation according to which being a mental even with a certain content has nothing to do with what the event is likely to cause. Then you might have a lot of trouble explaining the correlation. That is exactly the position that Fodor is in. On his account, the content of a mental representation type is determined by the kinds of thing that would cause tokens of the type to occur. Thus the content of a representation depends on its temporally backward-looking causal sensitivities. Those are evidently distinct from the representation's temporally forward-looking causal powers, what it, in turn, goes on to cause.

Fodor recognizes a version of the problem in Fodor (1994). His account of psychological content entails that it is referential. A person's 'Hesperus' representation and their 'Phosphorus' representations are causally sensitive to the same planet, hence, according, to Fodor, they have the same content. But, of course, the representations may behave differently in a person's psychological economy: they have relevantly different forward-looking causal powers. And so Fodor's account leaves open the possibility that psychology will have lots of predictive failures and be no good: for example, a Fodorian reference-

only psychology would predict that someone who wants to go to Hesperus and believes that boarding the *USS Morning Star* will get him to Phosphorus, would, ipso facto, be likely to board the *USS Morning Star*.

Fodor spends a lot of time trying to find mechanisms that keep referential psychology viable. These would be mechanisms that prevent Frege cases from happening too often. But he fails, for reasons that I will explain in chapter N.

So what could explain the correlation between a mental event's content and its causal powers? There are two kinds of explanation we might consider: metaphysically necessary and metaphysically contingent. Fodor looked for a metaphysically contingent one. And one would expect any account of content that relies entirely on historical facts to be in the same position and stand in need of supplementation with an account of contingent mechanisms that explain the correlation.

But obviously functionalism has a built-in explanation of the correlation, one that depends entirely on the metaphysics of content. What it is to be an event with a particular content is (in part) to have certain specific causal powers. So, for example, it is no accident that my desire to eat an Italian sausage tends to cause me to eat an Italian sausage, under certain conditions. Having that tendency is a metaphysically necessary condition for being a desire with that content. So, of course, any micro-properties that realise the desire will tend to cause events that realise my eating an Italian sausage.

Interpretationism is afloat in the same boat. According to interpretationism, a set of states of a physical system is representational iff there exists an interpretation of those states such that actual complex behaviour of the system consistently makes reasonable sense under that interpretation. The notion of interpretation involved is the logical one: there is a mapping from the states onto some suitable domain. According to interpretationism, the states then really represent what they represent in the logical sense, under any sensible interpretation.

Interpretationism can explain the correlation between a state's content and its effects. Being sensibly interpretable is a holistic property of a system. An interpretation of a given state will only make sense in the context of an appropriate interpretation of the state's causes and effects. And the relevant kind of appropriateness is just what is required to explain the correlation. For example. a given state will only be interpretable as, say, a desire to to eat an Italian sausage if, under certain circumstances, it causes a state interpretable as an eating of an Italian sausage.

Interpretationism is not usually thought of as a variety of functionalism. But (as leading interpretationist Robert Cummins admits (ref)) they are metaphysically equivalent. Whether a particular interpretation makes sense of a system depends only on causal relations among states of the system. So I shall use 'functionalism' to include interpretationism in its extension.

Functionalism's capacity to explain a correlation that cries out for explanation, while other accounts of content leave the mystery unexplained seems to be a strong consideration in favour of functionalism. I

shall proceed on the assumption that some variety of functionalism is true.

(6) Conclusion: epiphenomenalism versus overdetermination

If functionalism is true, then mental properties are dispositional, hence not causally efficacious. And epiphenomenalism is true. But mental properties are none the worse for that. They are just as real as paradigmatic physical properties like stiffness and strength. And they can play indispensable roles in excellent causal explanations, just like their physical counterparts. There is nothing wrong with or second-rate about dispositional properties.

I said above that I would present my view as a variety of epiphenomenalism. It seems an appropriate label. But notice that it is very similar to another view. Suppose we replace the notion of a causally efficacious property with, say, the notion of a casually explanatory property. We could explicate this notion by examples; the stone's stiffness and impact strength are causally explanatory in respect of its breaking the window, its colour and age are not. To be causally explanatory is to feature in a causal explanation in the familiar way. It is just that we now understand that that familiar has way a different character than we thought. We thought the properties were only nomologically related to the effects they explain, whereas actually they relate by metaphysical necessity. We thought the 'because' in the 'the stone broke the window because it was strong ...' was an *affect* because, whereas actually it is a *reason* because.

We might then use 'epiphenomenalism' to apply to the thesis that mental properties are not causally explanatory in respect of the effects of their possessors. Such terminological moves seem not unreasonable. We could then reasonably claim to be proponents of an overdetermination approach rather than of epiphenomenalism. Perhaps the difference between overdetermination and epiphenomenalism is largely rhetorical.

BIBLIOGRAPHY

Armstrong ..

Blakcburn

Fodor, J., 1974 "Speical Sciences

Gordon, J., *The New Science of Strong Materials or Why You Don't Fall Through the Floor*, Princeton University Press New Jersey 1968.

Holton
Kim, J., 1998 Mind in a Physical World, MIT Press, Cambridge USA.

Lewis 19xx functionalism

Lewis, D., 1986 On the Plurality of Worlds Blackwell Oxford UK and Cambridge USA