Danilo Šuster

Semifactuals and Epiphenomenalism

Draft (pre-print). Final version appeared in *Acta Analytica*, vol.16, Issue 26, 2001, 23-43
The issue was dedicated to the Symposium on Pierre Jacob: *What Minds can Do* (Maribor, December 15, 1997)

Abstract

Mental properties are said to be epiphenomenal because they do not pass the counterfactual test of causal relevance. Jacob (1996) adopts the defence of causal efficacy of mental properties developed by LePore and Loewer (1987). They claim that those who argue for epiphenomenalism of the mental place too strong a requirement on causal relevance, which excludes causally efficacious properties. Given a proper analysis of causal relevance, the causal efficacy of mental properties is saved. I defend the counterfactual test and epiphenomenalism of the mental against this critique. In causal counterfactuals we hold everything the same, take out the causal property and see if the effect property occurs. We do not replace the causal property with a barely different property as presupposed by LePore and Loewer. But, I recognize some general problems in making counterfactual claims about mental events, which raise doubts about the usefulness of the counterfactual test in general.

1.

In the age of materialism the problem of mental causation often appears as the problem of the epiphenomenalism of the mental. If every feature of mentality is "reducible" to material (physical) features then there is a threat that mentality *qua* mentality is doing no causal work. The mental *makes no difference* to the physical, it does not lead to behaviour that would not have happened in absence of the mental. The mental does not pass the *counterfactual* test of causal relevance. Let me introduce this test in terms of two examples.

> Meaningful sounds, if they occur at the right pitch and amplitude, can shatter glass, but the fact that these sounds have a meaning is irrelevant to their having this effect. The

> glass would shatter if the sounds meant something completely different, or if they meant nothing at all. Their having meaning does not help explain their effect on the glass. To know why the glass shattered you have to know something about the amplitude and frequency of these sounds, properties of the sound that are relevantly involved in its effect on the glass. [Dretske, 1989: 1-2]

In this passage, the epiphenomenal character of semantic properties of the sound for the shattering of the glass is expressed in terms of the conditional: “Even if the sounds meant something completely different, or if they meant nothing at all the glass would still shatter.” The amplitude and frequency of the sounds, on the other hand, are properties which are causally relevant for the shattering of the glass, because the following semifactual is *false*: “Even if the sounds had a different amplitude and different frequency, the glass would still shatter.” A semifactual is a counterfactual conditional with a true consequent and (factually) false antecedent. But now take the following passage from the book by Pierre Jacob in which he introduces the problem of mental causation:

> Suppose that my belief *c* that there is orange juice in the ice-box is a cause of my behaviour *e* (itself identified with my bodily motions), how could the semantic property of my belief *c* be causally efficacious in the production of *e* in the presence of the biological, chemical and physical properties of *c*? Does not the causal efficacy of the various biological, chemical, physical properties of my brain state token *c* preempt or screen off the causal efficacy of its semantic property? [Jacob, 1996: 216]

In this passage the problem of the epiphenomenalism of the mental appears in the form of the inefficacy of the content of intentional states. Causal inefficacy of mental properties, construed as semantic properties of one’s belief, is expressed in terms of the semifactual: “Even if my belief *c* did not have its actual semantic properties, it would still cause the same behavior *e* (*because of* the biological, chemical and physical properties of my brain state token *c*).”

More should be said about the mental entities we are talking about (events, states, properties, features, property instantiations ...) and their relation to the respective physical basis. But the general form of the counterfactual test of causal relevance seems to be clear. Suppose it is true both, “*c* is F” and “*e* is G”. “*c* is F” is then causally *ir*relevant for “*e* is G” if the following semifactual is true: even if “*c* is F” had been false, “*e* is G” would still have been true (because some other property of *c* is doing all the causal work required for *e*'s being *G*).

In this paper I will make two assumptions, accepted by the majority of authors in their discussion of the counterfactual test. I will assume that it makes sense to speak about causally relevant properties and I will assume that the conditionals of the form: "Even if an event *c* did not have its mental feature (property) *M,* it would still cause an event *e* to have feature (property) *B*" are meaningful and that they are sometimes true.

The first assumption is rather innocuous. Even if causation is primarily a relation between events it still makes sense to ask questions of the form: "What is it about events *c* and *e* that makes it the case that *c* is a cause of *e*?" and be able to answer them by saying "Because *c* is an event of kind *F* and *e* is one of kind *G*." So we say that the amount of explosive used in the explosion was causally relevant for the size of the crater (and not its shape, colour …). Causes have effects in virtue of their properties and not all properties of a cause are responsible for its effects. The second assumption is more controversial. It presupposes that it is possible that some event should have had exactly the same physical properties and different mental ones (or no mental properties at all). I will discuss this presupposition in the final section.

Jacob discusses the counterfactual test in connection with the issue of whether Davidson's anomalous monism (AM) entails the epiphenomenalism of mental properties. But the problem is more general, it is especially acute for non-reductive physicalism, a position favoured by Jacob. According to this position, physical entities and their mereological aggregates are all there is, but psychological properties (including content properties) are irreducibly distinct from the underlying physical and neurological properties and have their own causal powers. In this paper I will limit myself to the framework of anomalous monism discussed by Jacob. He defends Davidson against the charge of epiphenomenalism and adopts the strategy developed by LePore and Loewer (1987).

I will argue that these authors do not succeed in their defence of AM against the counterfactual argument for causal irrelevance of mental properties. I will start with a short explanation of AM and the charge of the epiphenomenalism of the mental as expressed by Sosa (1984) in his version of the counterfactual test. LePore and Loewer defend AM against Sosa by trivialising the criterion of causal irrelevance they ascribe to Sosa. They claim that the criterion is too strong since it classifies perfectly good cases of causal relevance as cases of causal irrelevance. I will try to show that this particular rebuttal of the counterfactual argument for epiphenomenalism is unsuccessful, because it is based on an erraneous criterion of evaluation of causal counterfactuals. Thus I will try to rehabilitate the threat to mentality posed by the counterfactual test. In the final section, I will address the problem of trivializing the counterfactual test of causal relevance in a way that is different from the objection raised by LePore and Loewer. Given supervenience, it is not possible that some event should have had exactly the same physical properties and different mental ones. I will argue that the issue is complicated and intertwined with the general problem of making counterfactual claims about the causal relevance of a certain feature. Still, my final verdict will be that the charge of epiphenomenalism has not been appropriately diverted.

2.

Davidson's anomalous monism (AM) results from three premises. (i) Propositional attitudes enter individual (or singular) causal relations (or interactions); they can be causes and effects of physical and other mental events. (ii) The principle of the strict nomological character of causation: every singular causal relation implies the existence of a strict physical law. In other words, individual causal interactions hold in virtue of some strict physical law. (iii) The principle of mental anomalism according to which there can be neither strict psychophysical nor strict psychological laws. From these premises Davidson infers AM, the view that (i) all events are physical or that every mental event must be token-identical to some physical event, and (ii) that not all events can be given a purely physical explanation.

Several authors have raised the problem of the *epiphenomenal* character of mental properties according to AM. Any individual causal relation requires the existence of a strict physical law but intentional psychological laws cannot be strict (physical) laws. So the worry is that anomalous monism makes mental features causally inefficient with respect to behavioural properties. This worry has been typically expressed by Sosa.

> A gun goes off, a shot is fired, and it kills someone. The loud noise is the shot. Thus, if the victim is killed by the shot, it's the loud noise that kills the victim. ... I extend my hand because of a certain neurological event. That event is my sudden desire to quench my thirst. Thus, if my grasping is caused by that neurological event, it's my sudden desire that caused my grasping.
>
> Yes in a certain sense the victim is killed by the loud noise; not by the loud noise as a loud noise, however, but only by the loud noise as a shot, or the like. Similarly, assuming the anomalism of the mental, though my extending my hand is, in a certain sense, caused by my sudden desire to quench my thirst, it is not caused by my desire qua desire but only by my desire qua neurological event of a certain sort. Beside the loudness of the shot has no causal relevance to the death of the victim: had the gun been equipped with a silencer, the shot would have killed the victim just the same. Similarly, the being a desire of my desire has no causal relevance to my extending my hand (if the mental is indeed anomalous): *if the event that is in fact my desire had not been my desire but had remained a neurological event of a certain sort, then it would have caused my extending my hand just the same* [Sosa, 1984: 277-78, italic is mine].

According to Sosa, neither the mentality of mental events nor the loudness of the loud shot are causally relevant to the respective effects. Let us take *M* to stand for a mental property of a certain mental event *c*, *N* for the underlying neurological (biological, physical) property of *c* and *B* for the certain behavioural property of an event *e*. From the text italicised LePore and Loewer extract the following schematic semifactual:

> If *c* were *N* but not *M*, then *e* would still be *B*.

Sosa's reasoning is supposed to show that only a mental state's physical property, not its mental (or semantic) property, can be causally efficacious in producing behavior. "c's being a certain neural state, *Nc*, screens off *c*'s being a desire to quench thirst, *Mc*, from *e*'s being an extending of the hand *Be*. More generally, ... neural properties screen off intentional properties."[LePore and Loewer 1987: 638]

Davidson's version of AM is not committed to the existence of properties and Sosa does not speak about properties. For Davidson, events are mental only as described, and the descriptions do not pick up the properties of these events. But the notion of a causally relevant property is firmly based and I will follow LePore and Loewer in considering the property version of Sosa's reasoning. They extract the following principle of causal relevance of the property *F* of the event *c* for the property *G* of the event *e* ("¬Fc > ¬Ge" stands for the counterfactual "If *Fc* had not been the case, then *Ge* would not have been the case"):

> *c*'s being *F* is causally relevant to *e*'s being *G*, iff (i) *c* causes *e*; (ii) *Fc* and *Ge*; (iii) ¬F*c* > ¬G*e*; (iv) *Fc* and *Ge* are logically and metaphysically independent; (v) There is no property *F** of *c* such that (*F*c* and ¬*Fc*) > *Ge* holds nonvacuously.

They distinguish between two types of causal relevance. The first type is backed up by a strong law, the second type supports counterfactual claims. But this distinction is not important for my discussion. I will also skip over clauses (i) – (iv) and concentrate on condition (v) which I will rewrite as a condition of *causal irrelevance*:

CI *c*'s being *F* is causally irrelevant to *e*'s being *G*, if there is a property *F** of *c* such that if *c* were *F** and not *F*, then e would still be *G*.

In particular,

CIm *c*'s being *M* is causally irrelevant to *e*'s being *B*, because there is a property *N* of *c* such that if *c* were *N* and not *M*, then *e* would still be *B*.

LePore and Loewer claim that CI as a condition of causal irrelevance is too strong. According to CI undoubtedly causally efficacious properties turn out to be causally irrelevant. Mental properties turn out to be epiphenomenal according to CIm, but CI has to be rejected and causal efficacy of mental properties is saved (given a proper analysis of causal relevance).

LePore and Loewer adopt the Lewis-Stalnaker approach in the analysis of counterfactual conditionals. "If it had been the case that A, then it would have been the case that B" is true if and only if B is true at all the worlds most similar to the actual world at which A is true (or A is true at no such world). LePore and Loewer also suppose that an event *e* that occurs at the actual world may occur or have counterparts that occur at others. When 'c' and 'e' are rigid designators of the actual cause event and effect event, we evaluate the conditional of the form "If *Fc* had not been the case, then *Ge* would not have been the case" in the most similar worlds to the actual world in which *c* exist and does not have *F* or *c* does not occur (have counterparts at all). The conditional is true just only in the case where these worlds are such that counterparts of *e* fail to have *G* or *e* fails to exist (have a counterpart). [LePore and Loewer, 1987: 636]

Consider now the mental event *c* and the behavioural event *e* in Sosa's example. *c* possesses some basic neurological property *N* and some mental property *M* (being a desire to quench the thirst), and *e* possesses property *B* (being a certain movement of the hand). The semifactual:

Sn If *c* were *N* but not *M*, then *e* would still be *B*.

is true iff *e* (or counterpart of *e*) is *B* at all the worlds most similar to the actual world at which it is true that *c* (or counterpart of *c*) is *N* but not *M*. In this case *N* screens off *M* from *B*, *M* is epiphenomenal. Consider now the symmetrical semifactual:

Sm If *c* were *M* but not *N*, then *e* would still be *B*.

According to LePore and Loewer this semifactual is also true. If *c* had been a desire to quench thirst, but had not had *N*, it would have some other neurological property *N**. In the closest possible world where *c* is *M* (a desire to quench thirst) but not *N*, it still causes *e* to have *B* (a certain movement of the hand). In general, semifactuals of the form Sm support the following claim:

CIn *c*'s being *N* is causally irrelevant to *e*'s being *B*, because there is a property *N** of *c* such that if *c* were *N** and not *N*, then *e* would still be *B*.

So it turns out, surprisingly, that neurological (physical) properties of mental events are causally irrelevant for behavioural properties according to CI. But why is *Sm* true? Jacob and many other friends of mental causation rely on the notion of *multiple realizability* [Jacob, 1996: 170]. The idea of mulitple realizability is that for any type of mental event there are many diverse ways in which it can be physically realized or instantiated or implemented. In the case of Sm – from the assumption of the multiple realizability of the mental by the physical we might argue that if *c* were not *N* but remained *M*, *c*'s being *M* would then be realized by some other neural property *N** of *c*. In this case, if *c* were *M* but not *N*, *e* would still be *B*. The mental would then screen off the physical.

This result – that mental screens off the physical – is unacceptable, so LePore and Loewer (Jacob follows them) claim that Sosa's reasoning relies on too strong a screening condition in general. According to Sn, an instance of CI, *N* screens off *M* from *B*, so *M* is causally irrelevant. But according to CI, Sm is true also, so by parity of reasoning *M* screens off *N* from *B*, so *N* is causally irrelevant too. Lepore and Loewer take this last consequence to be a *reductio* of CI. On their view, Sosa has assumed a sufficient condition for a property to be causally irrelevant (or to lack causal efficacy) which is too strong.

Not only is the symmetry between Sm and Sn supposed to be a refutation of Sosa's reasoning against AM, semifactuals of the type Sm are actually used in the arguments *for* the efficacy of the mental in general. Here is a quotation from Yablo (*m* stands for a mental event and *p* for a physical event):

> Then when do we attribute effects to mental causes? Only when we believe, I can only suppose rightly, that the effect is relatively insensitive to the finer details of *m*'s physical

> implementation. Having decided to push the button, I do so and the doorbell rings. Most people would say, and I agree, that my decision had the ringing as one of its effects. Of course, the decision had a physical determination *p*; but most people would also say, and I agree again, that it would still have been succeeded by the ringing, if it had occurred in a different physical way, that is, if its physical determination had been not *p* but some other physical event. And this is just to say that *p* was not required for the effect [Yablo, 1992: 278].

It seems that Yablo accepts the following instance of CIn:

> Yablo's decision having a physical determination *p* is causally irrelevant to the ringing of the doorbell, because there is a physical determination *p** such that if Yablo's decision were *p** and not *p*, then it would still have been succeeded by the ringing.

LePore and Loewer read CI in such a way that mental and physical properties turn out to be equally inefficient in the production of behaviour. From this they conclude that CI is too strong and they look for a weaker condition which would allow mental properties to be causally efficacious. Yablo seems to accept CI in terms of CIn as the *very* condition of the causal relevance of the mental. How about the symmetrical CIm which expresses causal irrelevance of mental properties? I think that Yablo would deny the nonvacuous truth of "If *c* were *N* but not *M*, then *e* would still be *B*." He would claim that it is impossible for *c* to be *N* but not *M*. I will comment on this exact reversal of Sosa's reasoning in the last section of this paper.

3.

Consider again the criterion of causal irrelevance, attributed to Sosa.

CI *c*'s being *F* is causally irrelevant to *e*'s being *G*, if there is a property *F** of *c* such that if *c* were *F** and not *F*, then e would still be *G*.

Is this really what Sosa (and other potential "epiphenomenalists") have in mind? For *any* property *F* of an event *c* one can construct a property *F** of *c* which screens off *F*. Take the event of throwing a stone at a bottle. The bottle shatters. We are inclined to say that the mass of the stone (and not the origin, age, texture … of the stone) is causally relevant for the occurrence of the effect. But wait, if the mass of the stone had been slightly different, then the bottle would still have shattered! Even if the mass of the stone had been very much different, but the throw was performed with greater force (or from a different angle), the bottle would still have shattered. Not only is CI too strong, it seems to be useless because *all* properties get screened off. According to CI there are no causally relevant properties whatsoever.

There is something plausible about Sm and the counterfactual test of causal relevance in general but CI does not capture this intuition at all. How do we test for the causal relevance of a certain feature?

According to Mill's method a casual claim involves a claim about the way things would have been without the cause. The idea is simple: keep everything the same, take out the cause event and see if the effect event occurs. This has been transformed into counterfactual analysis of causation. According to Lewis's analysis of causation between particular events, one event *c* is the cause of another different event *e*, when, whether an event *e* occurs or not causally depends on whether an event *c* occurs or not. Causal dependence between two events *c* and *e* is analysed in terms of the truth of two counterfactuals: O(c) > O(e) and ¬O(c) > ¬O(e), where "O(c)" represents the proposition that an event *c* has occurred.

We can generalize this approach to casual claims about properties: keep everything the same, take out the causally relevant property and see if the effect event still possesses the relevant property. Suppose an event *c* causes event *e*. Then it is true to say that if *c* had not occurred then *e* would not have occurred. This is true if and only if *e* does not occur in all the worlds *most similar* to the actual world at which *c* does not occur. Which non-*c* worlds are to be considered in the evaluation of the conditional "if an event *c* had not occurred, then an event *e* would not have occurred?" We might be inclined to say that the most similar non-*c* world is a world where a slightly different event *c** occurs. But then the effect *e* would have occurred as well! So *c* is not a cause after all? Or, in terms of properties, in many cases where some property (amplitude of the singing) is causally relevant to an effect (shattering of the glass), another closely related property (a slightly different amplitude) would have done the job as well. And if the event had lacked its actual property, it might well have had the closely related property. This objection was raised by David Braun [Braun D., 1995: 452] against the counterfactual test of casual relevance. A similar objection was made by Donald Nute (1980).

Nute claims that there are cases when one event depends causally upon another even though it does not depend counterfactually upon the other event. So *c* causes *e*, but it is false to say that "if an event *c* had not occurred, then an event *e* would not have occurred." Here is his example:

> Suppose we have a weight suspended by a cord from a rigid support. Let *c* be the cutting of the cord at a particular point and let *e* be the falling of the weight. If *c* and *e* both occur, then we should say that *c* causes *e*. But it is false that *e* would not have occurred if *c* had not occurred. I am assuming, of course, that if the cord had been cut at any point other than the point at which it was actually cut, then *c* would not have occurred. Furthermore, if the cord had been cut at any other point, the weight would still have fallen, i.e. *e* would still have occurred. Finally we simply need to observe that the worlds most similar to the actual world in which *c* does not occur would very likely be worlds in which the cord was still cut, but at a different point [Nute,1980: 95-96].

If this objection was decisive, then the counterfactual analysis of causation would be refuted. But what, exactly, is the antecedent of the causal counterfactual in question? "If the cord had not been cut (at all)…" or "if the cord had not been cut at *p* (but at some other point *p*')…?" Nute takes the second option and claims that this decision is supported by considerations of similarity between possible worlds. Is this really so? Suppose that a match is struck and it lights. What should we say about the counterfactual: "if the match had not been struck it would not have lit"? Well, if the match had not been struck the way it was struck but in a slightly different way (with a different force or angle or if the humidity was to some extent lower …) then the match would still have lit. So was the striking of the match not causally relevant for the lighting after all?

Something went wrong and it is easy to spot the difficulty. It has to do with the criteria of similarity. Here is Lewis's reply to a similar objection:

> ... a similarity theory needn't suppose that just any sort of similarity we can think of has nonzero weight. It is fair to discover the appropriate standards of similarity from the counterfactuals they make true, rather than vice versa. And we certainly do not want counterfactuals saying that if a certain event had not occurred, a barely different event would have taken its place. They sound false; and they would make trouble for a counterfactual analysis of causation not just here, but quite in general [Lewis, 1986: 211].

Nute's criteria of similarity would destroy any rationale for counterfactual analysis of causation. When you use counterfactuals to test for the casual relevance of a certain event, you remove that event and see if the effect event still occurs. You try to keep everything else the same, but of course, everything *cannot* be kept exactly the same. If the cord had not been cut then some facts (including, perhaps, facts about laws) would have been different. So, we consider situations which are as much like the original as possible and see if the effect occurs. What standards of similarity do we use? This is not an easy question. Lewis is prepared to accept intuitions about counterfactuals as our starting point. If we accept "if the cord had not been cut, then the weight would not have fallen" as true, then the worlds in which the cord is not cut and the weight does not fall are more similar to the actual world than the worlds in which the cord is not cut and the weight falls. Many would say that this is to put the cart before the horse. But here I need not discuss the problem of the analysis of counterfactuals. In order to respect the spirit of Mill's methods it is enough to acknowledge that when we remove the event – purported cause, we do *not* use standards of similarity according to which the original event is replaced by a barely different event!

Let us call the counterfactuals we use in testing our causal intuitions causal counterfactuals. Causal counterfactuals figure in the application of Mill's methods when we make claims about what would have been without the cause. But we use causal counterfactuals whenever we test for causal relevance of a certain feature (property, aspect …). So I will take Sosa's semifactual also to be a causal conterfactual. Standards of similarity used in the evaluation of casual counterfactuals have to be tailored according to our causal intuitions. Somebody like David Lewis, who analyses causation in terms of counterfactuals, would like to have uniform standards – principles we use in the assessment of causal counterfactuals are supposed to be general principles used in the assessment of any kind of counterfactual. This may or may not be the case – it has been noticed that the needs of a general account of counterfactuals and a general account of causation pull in different directions [Bunzl, 1984: 371]. But it seems clear that when *F* is causally relevant for *G*, then in the antecedent of the causal counterfactual "If *Fc* had not been the case …" we do not consider possible worlds where *c*, instead of *F*, possesses a slightly different property *F** (this point has also been made, in a different context, by Horgan, 1989: 60). But these seem to be exactly the standards of similarity used by LePore and Loewer!

In order to show that CI is too strong they invite us to consider the event of a hurricane, Donald, striking the coast and causing the streets to be flooded. That event is identical to the event of certain air and water molecules moving in various complex ways. The property of consisting molecules moving in such ways is *P*. The following counterfactual is then true: if hurricane Donald had not had property *P*, then it would still have caused the streets to be flooded. If hurricane Donald had not had property *P*, then a hurricane as much like Donald as possible, but without *P*, would have occurred. If hurricane Donald had not had property *P*, they claim, it would have had some property *P**, sufficiently similar to *P* and *P** events cause flooding. The hurricane would then have been a slightly different molecular event, but it would still have flooded the streets all the same. But, they claim, according to CI, it would be true to say that Donald's being a hurricane is causally irrelevant to its flooding the streets. This is absurd, so Sosa's criterion is not adequate [LePore and Loewer, 1987: 640].

Presumably LePore and Loewer identify the property of being a hurricane with the property *P*. And the most similar world in which Donald does not have the property *P* is a world in which it has the property *P**. So the following instance of CI is true:

> Donald's being *P* (i.e. being a hurricane) is causally irrelevant to flooding the streets, because there is a property *P** of Donald such that if Donald were *P** and not *P*, then it would still have flooded the streets.

According to CI the property of the event of being a hurricane would be causally irrelevant for the flooding of the streets. LePore and Loewer clearly presuppose that when we test for the causal relevance of the property *P* by entertaining the antecedent "If hurricane Donald had not had property *P* …" we replace *P* with a slightly different property *P**.

But this is not the way to evaluate *causal* counterfactuals. When we counterfactualize about the causal relevance of a certain feature we do not replace this feature with a slightly different feature. We remove the feature "totally". The intended reading of Sosa's semifactual is not:

> If the event that is in fact my desire had been a slightly different desire …

But rather:

> If the event that is in fact my desire had not been desire at *all* …

Let me roughly sketch the procedure of evaluation of causal counterfactuals. In the case of events, we remove the purported cause event, make the minimal changes in order to accommodate the removal and then see what happens. But we remove the event and all of its penumbra. We do not use Nute's criteria of similarity, so we do not consider the counterfactual:

> If the cord had not been cut at *p* (but at some other point p'), the weight would not have fallen.

When cutting the cord is a cause of the weight falling, we want to claim:

> If the cord had not been cut (at all), the weight would not have fallen.

I would like to argue that we use a similar procedure in the counterfactual test of causal relevance. Following the spirit of the previous criterion we remove the property, make the minimal changes in order to accommodate the removal and then see what happens. Very often the removal of a property leaves us with something impossible. Suppose that a bottle shatters because it is hit by a stone. We want to say:

> If the stone had no mass at all, the bottle would not have shattered.

But the object without a mass would cease to be the object. But neither do we want to say:

> If the stone had a (slightly) different mass, the bottle would not have shattered.

What we have in mind is something like:

> If the stone had a (considerably) different mass (and everything else remained the same as much as possible), the bottle would not have shattered.

One might argue that even a stone with a considerably different mass would no longer remain the same object. If we follow the suggestion by LePore and Loewer the evaluation of conditionals with such an antecedent is straightforward. Recall that according to them there are two ways for the statement of the type "It is not the case that *Fc*" to be true at a certain possible world. Either *c* exists at a possible world and does not have *F*, or *c* does not exist at this possible world at all. If we transfer this to the case of an object having a property: "If the stone had no mass at all, the bottle would not have shattered" is true because in the most similar possible worlds in which the stone does not exist, the bottle does not shatter.

If having a mass is a determinable, then having a specific mass is a determinate of this determinable. What counts as a considerably different determinate of a determinable is not fixed in advance. Sometimes we presuppose one answer and sometimes another. In order to activate the secret photosensitive mechanism which opens the tomb, Indiana Jones has to apply the exact force under the specific angle in the very narrow range of lighting conditions. In order to open the usual door, you just push the knob, a variety of forces and angles are allowed and light is irrelevant altogether.

When the property of a certain object *o* is epiphenomenal for a certain effect, then there is some *actual* property of *o* such that if the epiphenomenal property is removed or considerably changed, the remaining property would still do the required causal work. By an actual property I mean the property possessed by *o* in the actual world and retained by *o* in the counterfactual world (or worlds) in consideration. The very point of the counterfactual test of causal relevance would be lost if we took into account *non-actual* properties of *o*. So, in our case it is true to say:

> If the stone had no colour or a considerably different colour but still possessed its (actual) mass, the bottle would still have shattered.

It is easy to see that the example used by Dretske to show that semantic properties of the soprano voice are irrelevant for the shattering of the glass conforms to this pattern:

> Even if singing meant something completely different, or if singing meant *nothing at all* but retained the same frequency (pitch), the glass would still have shattered.

The same is true of the example used by Sosa:

> If the shot had no loudness or considerably different loudness (had the gun been equipped with a silencer) but retained the same (actual) ballistic properties, the shot would have killed the victim just the same.

In view of these examples, let me try to amend the criterion of causal irrelevance. I propose the following:

CI* *c*'s being *F* is causally irrelevant to *e*'s being *G*, if there is an actual property *F** of *c* such that if *c* did not posses a determinable of the kind *F* or possessed a considerably different determinate of the kind *F* but still remained *F** (and everything else remained the same as much as possible), then *e* would still be *G*.

I am not trying to specify necessary and sufficient conditions for causal irrelevance. I am simply claiming that the intuition of the counterfactual test of causal relevance is better captured by CI*. Let us apply CI* to mental properties. We begin with Sosa's semifactual:

> Even if the event that is in fact my desire had not been my desire (or it had had completely different mental properties) but had remained a neurological event of a certain sort, then it would have caused my extending my hand just the same.

If every event is token-identical to some physical event and causation requires strict nomological connections which are only obtained on the level of physical descriptions, this semifactual indeed seems to be true (given the initial assumption the antecedent of this conditional is meaningful). When we remove the mental property there is some remaining actual, neurological property, doing all the causal work.

Finally, let us consider the objection raised by LePore and Loewer to CI. Are there any undoubtedly causally relevant properties that turn out to be causally irrelevant according to CI*? Let us check the example of Donald the hurricane.

> Donald's being *P* (i.e. being a hurricane) is causally irrelevant to flooding the streets, because there is an actual property *P** of Donald such that if Donald did not posses a determinable of the kind *P* or possessed a considerably different determinate of the kind *P* but still remained *P**, then it would still have flooded the streets.

But of course, if Donald did not have *P*, then there would be no *actual* remaining property of Donald doing the causal work. There would be no Donald – the antecedent would then be true according to LePore and Loewer. But there would be no flood either, so the consequent would be false. Also, it seems at least very plausible to *deny* that if Donald possessed or consisted of a considerably different configuration of air and water molecules moving in various complex ways (*P**), then it would still have flooded the streets. It might or it might not. And the *denial* of the semifactual "if Donald did not posses a determinable *P* or possessed a considerably different determinate of *P* but still remained *P**, then it would still have flooded the streets" is according to CI* a necessary condition for the causal relevance of *P*. It might not be a sufficient condition, but here I am only interested in the comparison of CI and CI*. And CI* does not write off causally effective properties such as *P* as epiphenomenal just because in some possible world Donald possesses a slightly different property *P**.

How about the claim that if mental properties are epiphenomenal then so are the neurological? How do neurological properties pass the CI*?

> Even if the event that is in fact a neurological event of a certain sort had not been a neurological event (or it had had completely different neurological properties) but had remained my desire, then it would have caused me to extend my hand just the same.

Given our previous discussion, we are *not* supposed to consider the events of desire with slightly different neurological bases. Almost every contemporary philosopher will deny the possibility of an event having mental properties but *no* neurological properties at all. Given our laws of nature, the antecedent is impossible, there is no such event. Given the criteria used by LePore and Loewer the following conditional must be true:

> If the event that is in fact a neurological event of a certain sort had not been a neurological event (or it had had completely different neurological properties) but had remained my desire, then it would *not* be the case that I would have extended my hand.

Remember that "If *Nc* had not been the case, then *Be* would not have been the case" is true if in the most similar worlds to the actual world in which *c* does not occur (have counterparts at all) counterparts of *e* fail to have *B* or *e* fails to exist (have a counterpart). Clearly, in the worlds where there is absolutely no neurological event (no *c*), there is *no* behavioural effect either (no *e*). The semifactual "Even if the event that is in fact a neurological event of a certain sort had not been a neurological event (or it had had completely different neurological properties) but had remained my desire, then it would have caused my extending my hand just the same" is then *false* since not all of the closest antecedent worlds are such that the consequent is true in them. CI* does not write off neurological properties as epiphenomenal.

And how about an event that is in fact a neurological event of a certain sort but has completely different neurological properties? Of course, according to multiple realizability of the mental by the physical, any mental state is capable of implementation in diverse neural-biological structures in humans, reptiles, computers, Martians … . But recall the recipe for a causal counterfactual – keep the things as much like they were as possible, take out the causally relevant property and see if the effect occurs. In the counterfactual test we do not speculate about what would happen if the event remained Sosa's desire but Sosa was made from wires and silicon chips or had a completely different bio-neurological structure. Arguably, such an event would be very dissimilar to the original event. And we are interested in the question whether the very same behaviour would have been produced even if the event, which was in fact the event of Sosa's desire, had had completely different neurological properties. Again we are inclined to deny the possibility of such an event, so in the evaluation of the conditional we consider the most similar possible worlds in which the event with neurological properties does not occur. But then there is nothing left to do the causal work, so it would not be the case that Sosa would have extended his hand just the same.

Let me summarize – I think it is plausible to deny that the very same behavioural act would have been produced even if the event that produced it had had completely different neurological properties. So, according to CI*, neurological properties of an event are not screened off by mental properties of that event and the symmetry introduced by LePore and Loewer is broken.

I have tried to amend the condition of causal irrelevance in a way which incorporates insights about the evaluation of causal counterfactuals in the reflection on Sosa's semifactual. CI as a condition of causal relevance was based on the idea that in evaluating the counterfactual claims about a certain property we replace this property with a barely different property. But when testing for causal relevance of a certain feature we remove the feature and all of its "penumbra". The improved condition CI* better expresses our intuitions about causal relevance and according to CI* mental properties turn out to be epiphenomenal, but this is not the case with neurological properties.

4.

No doubt, a careful reader will notice some loose ends in my discussion. She might not agree with my evaluations of the critical conditionals and might be inclined to distribute truth values differently. I evaluated the antecedent "If the event that is in fact a neurological event of a certain sort had not been a neurological event (or it had had completely different neurological properties) but had remained my desire… " in the most similar worlds in which the event in question does *not* occur. But somebody might claim that the antecedent is impossible and has no literal interpretation at all. How could an event remain *my* desire and have no neurological properties or have completely different neurological properties? And it is equally impossible for the event that is in fact a neurological event of a certain sort to have completely different neurological properties and still *remain* my desire. A conditional with the impossible antecedent represents a misuse of a counterfacual construction. David Lewis proposed to classify all counterfactuals with an impossible antecedent as vacuously true. The only way to make *sense* of such an antecedent is precisely the procedure used by LePore and Loewer – the supposition that Sosa's desire is implemented in his brain in a (slightly) different way.

In this counterfactual scenario there would still remain some *actual* property doing the causal work. Namely, precisely the *mental* property of events being a desire. This is the way Yablo defends mental causation. We attribute effects to mental causes when we believe that the effect is relatively insensitive to the finer details of physical implementation of the mental event. One could even accept the amended CI* as the criterion of causal relevance and still claim that mental features are not screened off by physical ones. It could be argued that the only way to make sense of the counterfactual test is to accept Sm (mental causation) and deny Sn (neurological causation). Recall:

Sm If *c* (which is actually *M* and *N*) were *M* but not *N*, then e would still be *B*.

This semifactual is true when evaluated as suggested by LePore and Loewer. But symmetrical Sn does not make sense at all, even if we use the LePore and Loewer criteria of similarity:

Sn If *c* (which is actually *M* and *N*) were *N* but not *M*, then *e* would still be *B*.

How could a mental event retain its physical properties and lose or even slightly change its mental properties? Davidson himself argued that mental properties *supervene* on the physical. And nowadays the majority of philosophers would subscribe to strong psychophysical supervenience. Strong psychophysical supervenience is the claim that if something has a mental property at a time and in a world, then it has some physical property at that time in that world such that anything with that physical property, at any time and in any world, also has the mental property. So it is *impossible* that the event should have had exactly the same physical properties and different mental ones (or no mental properties at all). The counterfactual argument for causal irrelevance based on Sosa's semifactual is entirely beside the point, as noted by Zangwill [Zangwill, 1996: 78 –79]. If strong supervenience holds, the counterfactual asserted has an impossible antecedent. Given that an event has a physical property, it cannot fail to possess also its supervenient mental properties.

The argument for epihenomenalism of the mental was supposed to show that if we remove or radically change the neurological properties of a mental event, there will be no actual property of an event left which would be causally relevant for the behavioural effects. But if we remove or radically change the mental properties of a mental event, there will remain some actual, neurological properties which would still be causally relevant for the behavioural effects.

The opponent will claim that if we remove or radically change the neurological properties of a mental event, we destroy the event and the counterfactual test becomes useless. It is not possible to keep things as much like they were as possible, take out the causally relevant property and see if the effect occurs. Properties are lawfully connected and you can not isolate one without disturbing all the others. Counterfactual claims about a certain mental event which retains its mental property but loses its actual neurological property make sense only as claims about the event having slightly different neurological properties. It is possible (sensible) that some event should have had exactly the same mental properties and different physical ones. Moreover, assuming supervenience, it is *not* possible that some event should have had exactly the same physical properties and different mental ones (or no mental properties at all). So the symmetrical counterfactual claim about a certain mental event which retains its neurological property but no longer possesses its actual mental property does not make sense at all.

Is this a rehabilitation of LePore and Loewer (and Jacob)? The objection is based on the fact relating to the evaluation of counterfactuals. And this fact undermines the usefulness of the counterfactual test and counterfactual theory of causation in general.

Let me sketch a possible line of defence. In the frame of the counterfactual analysis of causation, developed by Lewis, the problem of *epiphenomena* can only be explained by supposing that some laws of nature connecting the cause event (causally relevant property) and the epiphenomenal event (property) are broken. Suppose that *c* (lowering of the air pressure) causes first *e* (barometer reading) and then *f* (storm), but *e* does not cause *f*. Suppose further, that given the laws of nature, *c* could not have failed to cause *e* and that, given the laws of nature and other circumstances, *f* could not have been caused otherwise than by *c*. It seems to follow that if the epiphenomenon *e* had not occurred, then its cause *c* would not have occurred and the further effect *f* of that same cause would not have occurred. So without *e* (changes in the barometer), there would have been no *f* and the storm turns out to have been caused by the changes in the barometer reading? Not so, according to Lewis. Rather, if there had been no changes in the barometer reading, changes in the air pressure would have occurred just the same and led to the storm [Lewis, 1993: 203].

Consider a setup described by Block [Block,1990: 147]. A metal rod connects a fire to a bomb. So long as the thermal conductivity of the rod is low, not enough heat is transferred from the fire to the bomb to cause an explosion. But if the thermal conductivity of the rod is increased enough (say, by altering its composition) then the heat from the fire will explode the bomb. There is a Widemann-Franz Law linking thermal and electrical conductivity under normal conditions (the same free electrons carry both charge and heat). In this setup, rising electrical conductivity together with other things being equal, is sufficient for an explosion. But electrical conductivity does not cause the explosion. Not in virtue of being a rise in electrical conductivity. Rather, the rising electrical conductivity is an inactive concomitant of the causally relevant rising thermal conductivity. It is the rising thermal conductivity that allows more heat to be conducted to the bomb, causing the bomb to explode. In this setup the rising electrical conductivity is epiphenomenal for the explosion, so we want to say:

> Even if the event, which was in fact the event of a rise in electrical conductivity, had not been the event of a rise in electrical conductivity but remained the event of rising thermal conductivity, the bomb would still have exploded.

But given the lawful connection, thermal conductivity can not be raised without raising the electrical conductivity. The antecedent is nomologically impossible – does this fact make counterfactual test useless? Not so, according to the criteria of similarity proposed by Lewis in his analysis of epiphenomena . Thermal conductivity would rise without the accompanying rise of the electrical conductivity. It is less of a departure from actuality to get rid of the rise in electrical conductivity by keeping the event of the rising thermal conductivity and the event of the explosion fixed and giving up some or other of the laws and circumstances in virtue of which the rising thermal conductivity could not have failed to be correlated with the rise in the electrical conductivity, rather than to hold those laws fixed and get rid of the electrical conductivity by abolishing the event of the rising thermal conductivity. Again, these criteria of similarity may be specific to causal conditionals, but such an account seems to be necessary for explaining the problem of epiphenomena.

This might also be a reply to the objection that Sosa's semifactual misfires, since, assuming supervenience, it is not possible that some event should have had exactly the same physical properties and different mental ones (or no mental properties at all). Suppose we understand supervenience of the mental on the physical as *superdupervenience* – ontological supervenience that is robustly explainable in a materialistically acceptable way [Horgan, 1993: 577]. The impossibility of an event having exactly the same physical properties and different mental ones (or no mental properties at all) would then be a certain nomological impossibility on a par with the impossibility of raising thermal conductivity without raising electrical conductivity. But we saw from Block's example that nomological impossibility does not deprive the counterfactual test of its discriminatory role in separating epiphenomenal and causally effective properties.

This rehabilitation of the counterfactual test might work if we take the antecedent in "if the event that is in fact my desire had not been my desire but had remained a neurological event of a certain sort …" to be nomologically impossible. There are, of course, other explanations of supervenience of the mental on the physical in terms of constitution and composition (mental states are constituted / composed by physical states). Would the metaphysical impossibility of an event having exactly the same physical properties and different mental ones (or no mental properties at all) deprive the counterfactual test of its discriminatory role in separating epihenomenal and causally effective properties? The answer is not so clear. It is notoriously difficult to make claims about the counterfactual identity of events. We have difficulties in assessing relatively simple matters – would my walking home be a different event if it happened a bit slower? We are even more at a loss when making counterfactual claims about the identity of mental events.

I think that we are confronted with a typical philosophical clash of intuitions – our opinions about causal counterfactuals pull in one direction whereas the generally accepted doctrine of supervenience pulls in the opposite direction. In the end this conflict might turn out to be decisive for the final verdict on the counterfactual test. Still, I believe that supervenience by itself does not solve the problem of the epiphenomenalism of the mental. But that is another topic.

REFERENCES

Block, N. 1990: "Can the Mind Change the World?", in G. Boolos (ed.) *Meaning and Method*, London: Methuen.

Braun, D. 1995: "Causally Relevant Properties", *Philosophical Perspectives* 9: 447-75.

Bunzl, M. 1984: "Causal Factuals", *Erkenntnis* 21: 367-384.

Dretske, F. 1989: "Reasons and Causes", *Philosophical Perspectives* 3: 1-15.

Heil J., Mele A. eds. 1993:, *Thinking Causes*, Oxford: Clarendon Press.

Horgan, T. 1989: "Mental Quasation", *Philosophical Perspectives* 3: 47-76.

Horgan, T. 1993: "From Supervenience to Superdupervenience: Meeting the Demands of a Material World", *Mind 102*: 555-586.

Jacob, P. 1996: *What Minds Can Do*, Cambridge: Cambridge University Press.

Kim, J. 1993a: "Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism", in Heil J., Mele A. eds. 1993: 19-26.

Kim, J. 1993b: *Supervenience and Mind*, Cambridge: Cambridge University Press.

LePore, E. and Loewer, B. 1987: "Mind Matters", *Journal of Philosophy* 84: 630-642.

Lewis, D. 1993: "Causation", in Sosa, E. and Tooley, M. (eds). *Causation*, Oxford: Oxford University Press, 193-204.

Lewis, D. 1986: *Philosophical Papers Vol.II,* Oxford: Oxford University Press.

Nute, D. 1980: *Topics in Conditional Logic*, Dordrecht: D. Reidel.

Sosa, E. 1984: "Mind-body Interaction and Supervenient Causation," *Midwest Studies in Philosophy* 9: 271-282.

Yablo, S. 1992: "Mental Causation", *The Philosophical Review* 101: 245-280.

Zangwill, N. 1996: "Good Old Supervenience: Mental Causation on the Cheap", *Synthese* 106: 67-101.