Counterfactual Overdetermination vs. the Causal Exclusion Problem

Georg Sparber
Department of Philosophy and Centre romand
for logic, history and philosophy of science
University of Lausanne
1015 Lausanne
Switzerland
Georg.Sparber@unil.ch

Abstract: This paper aims to show that a counterfactual approach to causation is not sufficient to provide a solution to the causal exclusion problem in the form of systematic overdetermination. Taking into account the truthmakers of causal counterfactuals provides a strong argument in favour of the identity of causes in situations of translevel causation.

The causal exclusion problem states a general consistence problem for the causal efficacy and/or distinctness of supervenient properties, such as, in particular, biological properties. To the extent that the subvenient basis is exclusively composed of tokens of physical properties the concept of supervenience in its strong, global form is commonly considered as the lowest common denominator of physicalism. Therefore, the general problem of causal exclusion applies in particular to the popular position of non-reductive physicalism.[1] In this paper I examine the claim that a counterfactual approach to causation can escape the inconsistency trap and save the non-reductive accounts of causally effective supervenient properties. For the sake of practical application I am restricting myself to a formulation of the problem focusing on biological properties. A non-reductive physicalist with respect to biological properties defends the following three propositions (Kim 2003: 151-152 and 157-158):[2]

1. (Sup) Biological properties supervene strongly on physical base properties. That is to say, if any system s instantiates a biological property B at t, there necessarily exists a physical property P, such that s instantiates P at t, and necessarily anything instantiating P at any time instantiates B at that time.

In more familiar terms (Sup) excludes the possibility of biological property variation without underlying physical property change. The principle includes an ontological

[1] Throughout the paper I use the concept of reduction in its ontological meaning and not in the epistemological sense, viz. the possibility to explain a theory's content in the framework of another. Non-reductive physicalism is therefore the thesis that macro-property tokens are ontologically distinct from arrangements of base-property tokens.

[2] I follow Jaegwon Kim in his presentation of the exclusion argument although I slightly change some aspects that are not of concern for our considerations and adapt the formulation for biological properties.

dependence clause between B-instances and P-instances. The second proposition concerns the irreducibility of the biological to the physical domain.

2. (Irr) Biological properties – types and tokens – are not reducible to, and are not identical with, physical properties.

The third proposition describes the causal status of biological properties as being efficacious, therefore vindicating together with (Irr) their treatment by an autonomous science that yields proper law-based explanations involving them.

3. (Eff) Biological properties have causal efficacy - that is, their instantiations can, and do, cause other properties, both biological and physical, to be instantiated.

The three propositions presented above are inconsistent when combined with the causal exclusion principle:

4. (Excl) No single event can have more than one sufficient cause occurring at any given time – unless it is a genuine case of overdetermination.

The relata of causal relations are events. Events are property instantiations at space-time regions. To be an event for a space-time region is therefore to instantiate some property. Every biological event that is a cause has a physical event as its base property instantiation (by (Sup)). Whenever there exists a biological cause, there exists a physical cause as well (namely the base property instance). The exclusion argument claims that in general, genuine overdetermination does not exist. Instead of two causes, there is only a single cause. In order to know which event has to be eliminated as a cause we have to resort to (Sup). Since in worlds like ours (governed by (Sup)) it is metaphysically impossible to delete the physical cause the supervenient one is to be eliminated, in our case the biological one. There remains only a decision to be taken about the way in which the biological event is eliminated as a separate cause: Either the alleged biological cause is not a second cause because it is identical with or reducible to a physical cause (non-Irr), or it is not a cause because it is causally inefficacious (non-Eff).
Instead of denying one of the three basic propositions accepted by the non-reductive physicalist, the overdetermination approach seeks to reject the consequence of the exclusion principle. Contrary to Kim's view that overdetermination is not a viable solution (in principle) an overdeterminist defends the position that in the case of biological causation we have regular overdetermination, which is explained by the ontological dependence in (Sup). There are two causes since there is a biological cause and the biological depends on the physical. The result is what may be called metaphysically dependent overdetermination. Barry Loewer outlines this possibility (Loewer 2001: 318-319). He claims that such an approach is only feasible if we adopt a counterfactual definition of causation, whereas a conception of causation as physical production cannot be saved from the inconsistency of the set of the four propositions above. It is the task of this paper to examine this alleged feasibility claimed by Loewer.

He could have the following in mind:[3] The relata of causation are events. Events are property instantiations. The event of the flower's having a certain gene that produces white petals and the physical base event of this event (among other things the instantiation of some DNA properties and further environmental physical conditions) are distinct since they are instantiations of different properties. The production of the flower's white petals supervenes on some change in the cells of the flower that is linked among other things to the fact that the cells reflect light waves in the specific manner as to appear white to our eyes. If this gene had not been there, the cells would not have reflected the light waves in this way. But it may be the case that if the specific physical base event (the DNA properties instances etc.) had not occurred, the flower could have reflected the light waves in the same way; for another physical base event could have occurred that would have been a sufficient condition for this way the cells reflect the light. There can be different true counterfactuals for the same effect. Hence, the counterfactual premises (the causes) cannot be one and the same thing. In what follows I examine the methodological deficiency as well as the ontological shortcomings included in the above reasoning. I finally claim that overdetermination is not a possible way out of the exclusion problem, not even on the basis of a counterfactual analysis of causation.

Concerning the method presented by the above argumentation the non-reductionist position seems to rely on the intuition that the abundant existence of true counterfactuals in our world makes possible the existence of distinct causal relations every time we use level distinct concepts to account for causes. By the fact that causal counterfactuals can differ while accounting for the same effect, the non-reductive physicalist seeks to show that the causes cannot be identical.

Following Loewer, the non-reductionist argument starts with a counterfactual theory of causation. Such a theory implies that causal counterfactuals can differ while being true and referring to the same effect. Second, the (Irr) principle says that biological properties – types as well as tokens – are not identical with physical properties. (Irr) implies that biological causes cannot be identical with physical causes. However, there is no implication from the claim of the difference of counterfactuals to the claim of the difference of properties (or their instances). The difference of causal counterfactuals could be due to different descriptions instead of different entities. I will argue that non-reductive physicalism cannot be saved from inconsistency by the claim that there can be different true counterfactuals referring to the same effect.

My argument follows a vertical direction from the ground to the top. Whether the irreducibility principle is justified or not has to be examined on the ontological level of the truthmakers of the corresponding causal propositions (in our case counterfactuals) on the one hand, and of the composition of property tokens on the other hand. Thus, my task is to show first that the truthmakers do not give us reasons to believe in two distinct causes and second that when looking at the property composition it is reasonable to believe in at least ontological token identity. Therefore overdetermination is disqualified as a possible solution for the causal exclusion problem.

Consider again the example situation:

[3] Loewer has not claimed the irreducibility of biological properties. His considerations concern exclusively mental properties. However, as the two cases are not distinct in principle they allow for analogous treatment.

1) This gene (biological token) causes the cells reflecting the light waves in a specific way (physical token) or formally: b causes p'.
2) This physical property token (among other things the DNA property tokens plus some environmental conditions) causes the cells reflecting the light waves in a specific way or formally: p causes p'.
3) The flower having this gene (the biological token b) supervenes on the base property token (p). Hence, there cannot be a difference in b without a difference in p, and b depends on p.

Assume a counterfactual definition of causation. In its most developed form a counterfactual definition refers to causation in terms of influence chains (Lewis 2000: 190). For example, b causes p' if and only if there is a chain of influence starting at b and finishing at p'. Influence is defined as a class of true counterfactual relations holding between alterations of the cause and effect events. An alteration is a small variation of an event. To have an influence from b to p' simply means that b and p' are counterfactually connected and that sometimes if one slightly changes b then one will slightly change p' as well. In other words: the relation of influence states that the cells' light reflection (p') depends on the gene (b) in the sense that some manipulations of the gene will have consequences on the way the cells reflect the light waves (and therefore on the petals' color). This is necessary and sufficient for a causal relation.
A counterfactual is true if and only if there is no world where its correspondent implication does not hold that is closer to our world than any world where the implication holds. For b and p' to be counterfactually connected means that worlds with b and p' as parts are more similar to our world than worlds with only b as part but without p'. To talk about worlds in this context has purely semantic purposes. We are not compelled to adopt an ontology of possible worlds (as Lewis does). The truth-values of counterfactuals supervene safely on events of the actual world as well as the laws and other aspects of similarity between worlds. It is exclusively the character of our world that accounts for the truth or falsity of a counterfactual statement. Possible worlds can therefore be regarded as mere constructions (most popularly of a linguistic kind) in order to clarify the meanings of counterfactual propositions (not to constitute them). In an analogous sense the alterations of an actual event are mere constructions in the form of equivalence classes having the actual alteration as starting point and basis. To the extent that the similarity aspects are primitive and objective we need not be bothered by the problem of a heavily multiplied ontology in the form of modal realism about possible worlds.[4] Nevertheless some problems about primitive modality remain unsolved (Lewis 1986b: 151).
The metaphysical motivation to promote counterfactual causation is its supervenience on intrinsic properties. Instantiations of intrinsic properties are independent of other properties' instantiations. Counterfactual dependence is not a basic property of our world. But the character of our world and the laws of nature determine the whole of counterfactual relations obtaining at our world. If counterfactual propositions express

[4] It follows from the definition of counterfactuals that their truth-values supervene on this-worldly facts if the respects of similarity are objective, constant and provide us with a sufficiently fine-grained order for possible worlds. The point-by-point construction of possible worlds entails a Humean view of the world without necessary connections between distinct point existences.

such relations between distinct events there is a causal relation between them. This position sharply contrasts with the view that causal relations are real relational features of our world. While counterfactual causation only states some sort of dependence among events, procedural causation defines the causal relation as a basic physical process where physical entities interact (Dowe 2000 and Salmon 1997 for example). Dowe and Salmon conceive causal relations as processes where entities exchange or transmit conserved physical quantities (as energy, momentum or charge). These conceptions accord with but are not entailed by the claim that among the basic properties of our world there are physical relational ones.

However, some philosophers reject the existence of any procedural basic property and defend an ontology with exclusively intrinsic base properties (Lewis 2000). One such metaphysic theory is Humean supervenience. It is the thesis that there are no relational base properties in our world apart from spatio-temporal relations, serving therefore the Humean view that no necessary connections between distinct existences (as events, for example) obtain in our world. The geometric arrangement of particular matters of fact (intrinsic properties instantiated at points in space-time) is sufficient to account for any property else instantiated (Lewis 1986a: ix-x). It is the task of ideal microphysics to provide us with an inventory of basic intrinsic properties. Candidates for this list are properties like having a mass, a charge and so one. Once there is an arrangement of particulars, every other property instance that obtains in our world is fixed as well. In particular counterfactual relations and biological properties as having a certain gene supervene on the arrangements of facts in our world.

What are the truthmakers of counterfactual propositions in the framework of Humean supervenience? Suppose that the relation R: b (the gene token) causes p' (the token of the way the cells reflect the light waves) supervenes on the arrangement X of particular facts. The extension of X is bigger than the simple union of event b and event p' as the evaluation of counterfactuals involves not only those events but also aspects of factual and nomological similarity. Suppose further that the relation R': p causes p' supervenes on the arrangement Y of particular facts. By (Sup) we know that the extension of Y is at least as big as the one of X. There are reasons to believe in identical truthmakers for propositions expressing R and R', namely reasons I will present when it comes to the question how supervenient property tokens are composed and what they really are.

One truthmaker admits of an infinity of different propositions all of which are made true by it. But in this case their difference does not reflect an underlying ontological difference in relations. Their difference is only a descriptive one. They talk in different manners about the same things. Until now we do not have reasons to believe in the identity of the two propositions containing the relations R and R'. Nor do we have reasons to believe in a difference in truthmakers for those propositions. What we know for the moment is only that there is a logical covariation in the sense of (Sup) between the instances of a supervenient property B and its subvenient base property P. Furthermore, there is a one-directional ontological dependence of every B-instance on some P-instance.

To answer the question of whether or not counterfactual overdetermination is possible let us consider the core statement of the non-reductionist argument, which asserts that the following is possible: this gene (token) could have had a different subvenient base property instance. In other words: this specific gene (token) can stay the same while its base property instance (the p token) changes. This would imply that the causal relations R

and R' involving the events b (gene token) and p or p' (physical property tokens) respectively are different and that their difference is based on different truthmakers.
There are several considerations that conflict with the claim that the same property token can have distinct physical base tokens. The first problem consists in the formal impossibility for an actual event to be strictly identical with a non-actual possible event. In terms of possible worlds, two events, as property instantiations at space-time regions belonging to different possible worlds, cannot be identical. And even if they belonged to the same possible world, they would still be distinguishable with respect to coordinates proper to the instantiating space-time region. The talk of possible worlds in this context can again be considered as purely semantic and non-compelling towards ontological commitments.
The second problem concerns the composition of supervening property instantiations within the scope of physicalism. The minimal necessary and sufficient condition for being a physicalist is commonly identified with the acceptance of the following global supervenience thesis: A minimal physical copy of the actual world is a copy simpliciter (Jackson 1998: 12). Every physicalist faces therefore an exclusively binary choice: Either the property instances of a supervenient property are exhaustively composed of physical base properties (in Lewis' case they are intrinsic and instantiated at point-regions) or such instances are composed of physical properties plus something more. The first alternative is known as the token identity conception whereas the second is called emergentism. The initial argument of the non-reductive physicalists claims that emergentism combined with counterfactual causation can save the (Irr) principle from token identity. Emergentism faces serious troubles when it is in need of explanation where the additional features of the supervenient property instances come from. Its essential claim is that they are not uniquely founded in physical property arrangements but that there is something extra, i.e. the emergent feature. Suppose with the emergentists that this feature permits to individuate supervenient macro-properties. Their tokens can stay the same while the correspondent physical base tokens change because the qualitative emergent feature stays the same. But even if this feature is regarded as the essence of macro-property tokens, they are also composed of basic physical property tokens. Identity of emergent features is not a sufficient criterion for the identity of macro-property tokens. Some of their components will change when the physical base token changes. If a difference in composition between the first gene token (b supervening on p) and the second one (b* supervening on an alternative physical base p*) can be found, they cannot be the identical, although they may share the same emergent qualitative features. In physicalistic emergentism the gene tokens and their correspondent physical base tokens are therefore coordinated in the following symmetric sense: there cannot be a difference in b without a difference in p by (Sup), nor can there be a difference in p without a difference in b because different composition entails distinctness. These considerations apply only to events as property tokens instantiated at space-time regions. Substances as animals or plants can stay the same although their composition changes because they are extended in time. Token events cannot do so, because they lack temporal parts. An event token whose composition changes becomes another token albeit of the same type. [5]

[5] There are non-reductionist attempts to consider event token identity not only as compositional identity but also as identity of modal and essential features. Since higher-level tokens can behave for example modally different from their corresponding base level cluster tokens they cannot be identical with the second. While

The truthmakers of causal counterfactuals are composed of causes, effects and some respects of factual and nomological similarity. The non-reductionists' claim that the truthmakers of translevel causal counterfactuals are different from their corresponding truthmakers of base level counterfactuals because the latter can vary without changing the former has been shown to fail. The distinctness criterion of macro-property tokens vanishes. One can no longer stipulate their ontological difference for the alleged reason of different counterfactual behaviour. Since biological events cannot causally behave in any way that permits to distinguish them from the underlying physical causes, there is no reason to believe in overdetermination. In other words: That the truthmakers of causal counterfactuals are not distinguishable is a strong reason to believe in their identity. In that case the difference in counterfactuals does not reflect any underlying ontological difference of truthmakers but it is only a difference in description referring to the same entities.

Since the main criteria to differentiate between the truthmakers of causal counterfactuals cannot be applied, the fact that one cannot distinguish them in this manner is sufficient to turn towards the second possibility of how macro-property tokens are composed, namely token identity. According to the token identity conception every supervenient macro-property instance is identical with an arrangement of subvenient micro-property instantiations. This identity explains the above stated logical covariation: because they are identical a causally effective macro-property instance cannot change without a change in its subvenient base property and if the subvenient base property changes, the supervenient property instance has to change as well. We have therefore identity not only concerning the tokens b and p but also identity of the relational property of causality: there is one and the same cause, one and the same effect and the definition of how to evaluate a counterfactual statement does not change. From identity of the relata and identity of the counterfactual relation follows identity of causal relations. We do not even face the question of overdetermination since we do not have two causes anymore.[6] All we have is two different descriptions of the same ontological entity referred to as the causal relation. A physicalist should get used to the fact that special sciences do not talk about other things than physics, but only talk differently about the same things.

---

the compositional coincidence satisfies the causal exclusion requirement of non-competing causes, the difference in modal characteristics makes for the explanatory autonomy and the irreducibility of higher-level causal tokens and their causal powers. Such positions defended by Derk Pereboom (Pereboom 2002) and Stephen Yablo (Yablo 1992) imply realism concerning modal properties (or essential properties for Yablo) that make for the identity of causal powers of tokens. But it is at least questionable if those features are identity criteria for property tokens instead of rather being definitional (by means of the functional role) for property types. While the first position is an attempt of using the multiple realization argument to motivate token irreducibility, the second claim is perfectly coherent in joining causal homogeneity of macro types with token identity. The multiple realization of the Theseus ship for example cannot be taken as a reason for one of its realizers to be non-identical with the ship itself, for when being realized the Theseus ship is a property type and not a token. Nevertheless one specific ship instance might well be identical with a physical cluster token of its composing elements. Compositional regards are then decisive for the identity or distinctness of tokens while modal or essential characteristics are decisif for the identity or distinctness of types.

[6] It is sufficient to have token identity to exclude overdetermination, but token identity does not imply type identity for well-known reasons of multiple realization. Therefore my result does not necessarily contradict the irreducibility principle for property types as formulated by Kim.

The supposition of token identity tells something in addition to the principles as stated by Kim in the course of the presentation of his causal exclusion argument. It is not included in the supervenience thesis and conflicts with the irreducibility principle that tries to establish a certain autonomy of supervenient properties in relation to subvenient properties. For Kim token identity reflects the more general fact of type identity since "the relevant sense in which an instance of M [a mental property in particular but any supervenient property in general] = an instance of P requires either property identity M = P or some form of reductive relationship between them" (Kim 2003: 157). Kim adopts the view that different properties have to reflect their distinctness by different causal powers (Kim 1998: 103). Or equivalently, no difference in causation implies there be no difference in property. As our examination of truthmakers of counterfactual causal statements shows, we have strong reasons to believe in the identity of tokens in the case of translevel causal situations. But in his argumentation Kim gives little evidence to believe in type identity. He does not explain why he thinks that token identity in causal situations engages us to accept the more general type identity. Therefore I do not follow him in his conclusion without further examination that goes beyond the scope of this paper.

What I claim is only that in regarding the truthmakers of causal counterfactuals in situations of alleged overdetermination there are no good reasons for believing in there being two distinct causes. I take this as an argument to believe in their identity. I conclude that a counterfactual theory of causation cannot escape the inconsistency trap affecting the set of premises that defines non-reductive physicalism by taking rescue to overdetermination for macro-property causation. An overdeterminist solution to the causal exclusion problem does not simply follow from the proper character of a counterfactual theory of causation as originally claimed by Loewer. To correct his alleged statement we can say now: The relata of causation are events. Events are property instantiations. It is not possible that the cause event of having this gene stays invariant while its physical base event changes. The event of the flower's having a certain gene and the physical base event are therefore not distinct. If that gene token had not been there the cells would not have reflected the light waves in this specific way. And if this physical DNA properties and further conditions had not been instantiated the cells would not have reflected the light waves in this way. These two counterfactuals are only distinct in the way they describe the same causal relation. In his reasoning on causal exclusion Kim leaves aside any discussion of what macro-property instances really are in favour of his famous exclusion principle to get to the result that causally effective supervenient properties are identical (or reducible) to subvenient base properties. Kim's claim that token identity entails type identity, as far as causally effective properties are concerned, allows for criticism. His exclusion argument refutes the emergentist attempt to save supervenient causation and my consideration of how properties are constituted extends his refutation to the case where such a theory is combined with a counterfactual theory of causation. In general, counterfactual causation cannot provide by itself a solution for problems about causal exclusion or overdetermination. It is the exclusive task of the domain of property theories to answer such questions no matter whether one conceives causal relations as physically productive or whether simply as counterfactual dependences. In this sense my thesis can also be considered as an attempt to rehabilitate

counterfactual theories of causation by countering the "anything goes" prejudice commonly addressed to those approaches.

References:

Dowe P., 2000, *Physical Causation*, New York: Cambridge University Press.

Jackson F., 1998, *From Metaphysics to Ethics*, Oxford: Clarendon Press.

Kim J., 1998, *Mind in a physical World. An essay on the mind-body Problem and mental Causation*, Cambridge: MIT Press.

Kim J., 2003, 'Blocking Causal Drainage and Other Maintenance Chores with Mental Causation', *Philosophy and Phenomenological Research*, LXVII, No. 1: 151-176. Reprinted in Kim J., 2005, *Physicalism, or something near enough*, Princeton: Princeton University Press.

Lewis D., 1986a, *Philosophical Papers. Volume 2*, Oxford: Oxford University Press.

Lewis D., 1986b, *On the Plurality of Worlds*, Oxford: Blackwell.

Lewis D., 2000, 'Causation as Influence', *The Journal of Philosophy*, 70: 182-197.

Loewer B., 2001, 'Review of Jaegwon Kim, Mind in a Physical World', *The Journal of Philosophy,* 98: 315-324.

Pereboom D., 2002, 'Robust Nonreductive Physicalism', *The Journal of Philosophy,* 99: 499-531.

Salmon W., 1997, 'Causality and Explanation: A Reply to two Critics', *Philosophy of Science,* 64: 461-477.

Yablo S., 1992, 'Mental Causation', *The Philosophical Review,* 101, No. 2: 245-280.