

# On Horgan's Causal Compatibilism

Janez Bregant, Maribor

## Introduction

It is quite obvious why the antireductionist picture of mental causation, which rests on supervenience, is such an attractive theory. On one side it secures the mental a preservation of its unique and different nature; on the other side it tries to place the mental in our world in a way that is compatible with the physicalistic view.

But Kim's argument from supervenience reminds us that while trying to do so the antireductionists face the following dilemma: if mental properties have causal powers we risk a violation of the causal closure of the physics and threaten our physicalistic position; if they do not have them we adopt epiphenomenalism, which denies the mental causal powers of any sort. So, either we violate the causal closure of the physics or we adopt epiphenomenalism. We can write it as an argument in the form of a constructive dilemma and may call it *the causal dilemma*:

- (1) Either mental properties have causal powers or they do not.
- (2) If they have them we violate the causal closure of the physics.
- (3) If they do not have them we adopt epiphenomenalism.

---

∴ (4) Either we violate the causal closure of the physics or adopt epiphenomenalism.

It is clear that for the antireductionists who want to preserve physical monism and mental realism none of the horns represents a true alternative and that the causal dilemma should be avoided at all costs.

The first part of this article describes Horgan's antireductionist answer to the causal dilemma, the view that he calls causal compatibilism, which is supposed to save mental causation. The second part introduces my objection to Horgan's solution based on Kim's distinction between micro-properties and micro-based properties showing that causal compatibilism leads any antireductionist to unbearable consequences, which leaves the causal dilemma unsolved.

## 1. Horgan's Answer to the Causal Dilemma

Horgan's answer to the causal dilemma consists of three ideas that together represent a non-reductivist solution of mental causation. Let's explain the first one by the help of the following: suppose that my desire for water (event  $a_1$ ) causes the drinking of it (event  $b_1$ ). Such a causal explanation emphasises the dependence of the effect  $b_1$ , described as a token of type  $b$  events, on a cause  $a_1$ , described as a token of type  $a$  events. Usually such a dependence requires for  $a_1$  and  $b_1$  to be a part of a counterfactual dependence, that directly connects a types (desires for water) with  $b$  types (drinking of water) in the sense that if  $a$  appears  $b$  appears also and if  $a$  does not appear  $b$  does not appear either, which could be symbolised as  $a \square \rightarrow b$  and  $\neg a \square \rightarrow \neg b$ . But in Horgan's opinion this is not enough for a causal explanation. He thinks that a similar but much more general counterfactual dependence connecting desires and drinking (or beliefs and

drinking, or intentions and drinking, or sensations and drinking, in short, whole groups of mental phenomena and actions) must exist, and he calls it a *pattern of counterfactual dependence*. His first idea is then the claim that we must, when analysing the causal relevance of properties, take into account also their capacity for being part of such counterfactual relations.

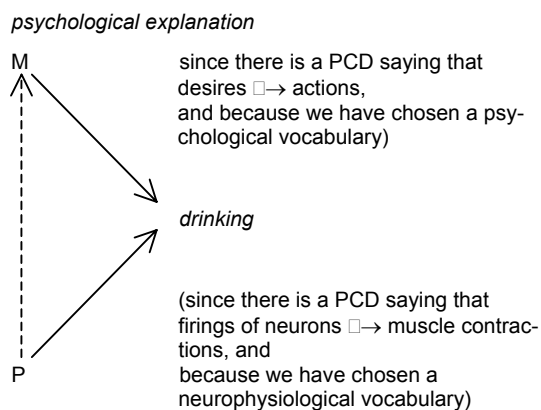
The second idea constituting his picture of mental causation is hidden in the following passage: "Which kinds of dependence patterns are most typical will be a context-relative matter, governed largely by the interests of those doing the explaining and inquiring." (Horgan 2001) One crucial feature of context which must be taken into account, when ascribing the causal efficacy to properties is that it involves factors and standards. The former determines the latter in the sense that, for example, to say when it is cold outside depends on being Slovene or Eskimo. The choice of vocabulary with which we want to describe a causal chain is one of the factors affecting a change of a context. Horgan calls them *context parameters* and thinks that they have a crucial role in event causation because they determine at which level of causal explanations we are and what properties we mention as causes. He says: "If we pose our questions and offer our answers in psychological vocabulary, for instance, then normally the relevant patterns of counterfactual dependence will be ones involving psychological properties /.../." (Horgan 2001) His second idea is then the claim that the notion of causation depends on parameters; by changing them we change the context; and by changing the context we change a type of causal explanation

Horgan's third idea composing his solution of the causal dilemma is based on the belief that there are different patterns of counterfactual dependence describing the causal activity of the same phenomenon in different vocabularies. The vocabulary that we choose simply depends on the ontological level at which we are and on the needs that it has. A physical effect, drinking for example, could be explained in neurophysiological terms as a consequence of neuron firings, or in psychological terms as a consequence of thirst. It is more than interesting that in his opinion such different causal explanations do not exclude each other but are, since they represent robust patterns of counterfactual dependence at different ontological levels, compatible.

The three ideas combined give us a picture of causation according to which the same phenomena can be parts of different patterns of counterfactual dependence; they appear at different ontological levels and it is the context that determines which of them is the most appropriate for the causal explanation. But sometimes the same phenomenon can be described in terms of different ontological levels producing different causal explanations that do not exclude each other. Their compatibility in Horgan's opinion originates in supervenient relations presumably holding among *different* ontological levels. According to them psychological patterns depend on lower located physical patterns. He writes: ".../ the mental properties that are causal properties at the psychological level have their causal efficacy via the causal efficacy of physical causal properties [that are causal properties at the physical level] that realize them." (Horgan 2001) Therefore causal powers

of mental properties are not over and above causal powers of physical properties, and this is why the causal closure of the physical domain is not violated.

Horgan's solution of mental causation could be illustrated by the following diagram (suppose again that my desire for water *M* causes *drinking*; note that *PCD* stands for the pattern of counterfactual dependence):



(The broken arrow denotes a relation of supervenience; the solid arrows designate a relation of causation.)

This looks very much like a case of causal overdetermination but Horgan denies this by saying that it is true that "given the specific level of description that is contextually appropriate for causal explanation, several properties are co-instantiated, all at the relevant level of description, each of which is such that its instantiation is independently causally sufficient (in the circumstances) for the effect." (Horgan 2001) But since causal overdetermination is something that arises when two causes appear at the same time at the *same* level mental, properties and their realizers occurring in different patterns of counterfactual dependence holding at different ontological levels do not face this problem.

It seems that such a solution of mental causation, called *causal compatibilism*, secures the mental a causal status. It supposes the existence of real causation at different ontological levels and says that despite the causal closure principle, causal explanations of the physical level are not incompatible with causal explanations of the psychological level.

## 2. Objection to Horgan's Causal Compatibilism

Does Horgan's answer really avoid the dilemma posed by Kim? Fortunately, Horgan does not underestimate metaphysics, as some philosophers suggest; on the contrary, he believes that the compatibility of causal explanations at different levels rests on supervenience. However, the idea that supervenience can secure the mental the causal efficacy via the physical without the worry of causal overdetermination turns out to be highly problematic and makes his answer controversial. Why?

Horgan says that mental properties are causally efficacious via physical properties and therefore there is no danger that the mental, while causing the physical, violates the causal closure principle. But isn't this something that the supervenient model of causation supposes in the first place in order to be a real non-reductivist alternative to

other solutions? Of course. The supervenient model of causation claims that a mental property (*M*) supervenes on a physical property (*P*) and that, for example, drinking supervenes on a muscle contraction; it further says that *M* causes drinking via *P*, which causes the relevant muscle contraction, and that this is enough to avoid a violation of the causal closure principle. But Horgan is aware that this is nothing new and that the real challenge is yet to come, namely, how to deal with the problems accompanying the supervenient model of causation: a critique that even on this model we must still deal with two different and sufficient causes of drinking, *M* and *P*. This makes the claim that the supervenient model of causation avoids the violation of the causal closure of the physics highly doubtful.

However, Horgan thinks that we can very well avoid causal overdetermination; his line of thought is represented by the following passage: "i.e., even after we contextually fix the operative score in the causal-explanation game in a way that restricts the relevant patterns of counterfactual dependence to those involving properties at a specific level in the hierarchy of the sciences, we still find several properties instantiated that each figure in the contextually relevant dependence patterns in a way that makes each property-instantiation an independently causally sufficient condition for the effect." (Horgan 2001) Again, the reason that we do not end up with causal overdetermination is, in his opinion, the fact that mental properties and their physical realizers appear in different patterns of counterfactual dependence at different ontological levels, whereas causal overdetermination always emerges when two causes occur at the same time, at the same level.

And why is this problematic? Well, one might say that Horgan's belief about the compatibility of such different causal explanations involving different causes, which rests on supervenient relations that presumably hold among properties of different levels, is simply false. This appeals thereby to Kim's distinction between microproperties and micro-based properties proving that supervenience is a relation that holds among properties of the same level and not among properties of different levels. But before we get to his idea, something else must be said. It is widely accepted that physical realization involves a claim which says that macro or second-order properties (a macro-property is for example a colour) are realized by first-order or microproperties (a microproperty is for example a spinning of quarks). Moreover, the understanding of the notion of a second-order property, to which we usually appeal states the following: "For something to have a second-order property is for it to have one or another of its realizers, that is, a first-order property satisfying the specification that defines the second-order property." (Kim 2001)

But this is bad news for Horgan since according to this we have, when going to the kitchen to get a glass of water, both properties: we are thirsty and the neurons that realize our thirst are firing. It seems that in the layered model of the world a second-order property and its realizer appear at the *same* level and that they are properties of the same object. Therefore, while speaking about second-order properties and their realizers we do not move either upward or downward but always stay at the same level. Kim says: "In general, supervenient properties and their base properties are instantiated by the same objects and hence are on the same level" and he continues »so microphysical, or mereological, supervenience does not track the micro-macro hierarchy." (Kim 2001)

The reason for the mistaken belief that higher-order properties supervene on microproperties, which may make us think that supervenience is a relation among properties of different levels, lies in a too strict understanding of supervenience. We become confused because realizers are usually described in terms of their microstructure, which leads us to the false belief that they are microproperties appearing at lower levels. But in fact realizers belong to higher levels, merely their description is different from those of other higher-order properties, and therefore also represent macroproperties. Transparency, for instance, is a macroproperty which can be realized only by molecules having the right combination of atoms, i.e. water. But its realizer is, despite being described as  $H_2O$ , a macroproperty, or at best *micro-based property*, since it is a group of atoms that have a property being transparent only if formed in this special way.

We could speak about real microproperties realizing macroproperties and about changing levels *only* if transparency were realized by its single atoms, or going even lower, by parts constituting them. But this can't happen since transparency is a property that can be realized only by molecules with the right structure. It is a feature of molecules and not atoms or quarks and it can appear only at the molecular level, which makes it impossible for its realizer to belong to any other level but the molecular one. So, when we say that Peter's desire (*M*) supervenes on his microphysical properties (*P*) we think only of its dependence on his micro-based properties (*not* microproperties) and relations holding among them. *P* on which *M* supervenes is therefore "the property of having such-and-such proper parts that have such-and-such properties and are configured by such-and-such relations." (Kim 2001) The property that they have as a whole, e.g. to be *P*, does not belong to any of its parts, just like the property that two atoms of hydrogen and one atom of oxygen have as a whole, e.g. to be  $H_2O$ , does not belong to any of its single atoms.

Supervenience is therefore a relation that holds at the *same* level and since causal overdetermination also emerges at the same level it looks like we are only a step away from excluding the mental from causal explanations.

The dilemma that Horgan faces now arguing for causal compatibilism is the following: either keep claiming that the compatibility of different causal explanations rests on supervenient relations that hold among them, which leads to causal overdetermination; or give up supervenient relations as something that can secure such compatibility and thereby risk having no metaphysical ground on which to base his solution of mental causation, regardless that this seems to be necessary for it. In the standard form, the dilemma could be written as follows:

- (1) Either we accept supervenience as a relation securing the compatibility of different causal explanations or we give it up.
- (2) If we accept it we end up with the causal exclusion of the mental (since supervenience as well as causal overdetermination are features of the same level supervenience can't make different causal explanations compatible with each other).
- (3) If we give it up we threaten our physicalistic view (because we do not meet a required condition for physicalism).

∴ (4) Either we end up with the causal exclusion of the mental or threaten our physicalistic view.

Since none of the horns is acceptable for the antireductionist like Horgan we might conclude that his solution of mental causation which supposes the compatibility of different causal explanations of the same phenomenon on the basis of supervenient relations holding among properties of different levels, does not represent a real non-reductive alternative, which leaves Kim's causal dilemma, at least from the antireductionist point of view, still unsolved.

## Literature

- Horgan, T. 2001 "Causal Compatibilism and the Exclusion Problem", *Theoria*, 40, 95–116.  
 Kim, J. 2001 *Mind in a Physical World*, Cambridge: MIT Press.