

Exclusion, Still Not Tracted

Douglas Keaton
Department of Humanities
Flagler College

Thomas W. Polger
Department of Philosophy
University of Cincinnati

Abstract

Karen Bennett has recently articulated and defended a “compatibilist” solution to the causal exclusion problem. Bennett’s solution works by rejecting the Exclusion principle on the grounds that even though physical realizers are distinct from the mental states or properties that they realize, they necessarily co-occur such that they fail to satisfy standard accounts of causal over-determination. This is the case, Bennett argues, because the causal background conditions for core realizers being sufficient causes of their effects are identical to the “surround” conditions with which the core realizers are metaphysically sufficient the states or properties that they realize. Here we demonstrate that the background conditions for the causal sufficiency of core realizers for their effects are not identical to the core realizer’s surrounds, nor do backgrounds necessitate such surround conditions. If compatibilist solutions to exclusion can be defended, a different argument will be needed.

Keywords

exclusion argument; core realizer; total realizer; Karen Bennett

Word Count

6828

Exclusion, Still Not Tracted

Abstract

Karen Bennett has recently articulated and defended a “compatibilist” solution to the causal exclusion problem. Bennett’s solution works by rejecting the Exclusion principle on the grounds that even though physical realizers are distinct from the mental states or properties that they realize, they necessarily co-occur such that they fail to satisfy standard accounts of causal over-determination. This is the case, Bennett argues, because the causal background conditions for core realizers being sufficient causes of their effects are identical to the “surround” conditions with which the core realizers are metaphysically sufficient the states or properties that they realize. Here we demonstrate that the background conditions for the causal sufficiency of core realizers for their effects are not identical to the core realizer’s surrounds, nor do backgrounds necessitate such surround conditions. If compatibilist solutions to exclusion can be defended, a different argument will be needed.

Keywords

exclusion argument; core realizer; total realizer; Karen Bennett

Word Count

6828

Exclusion, Still Not Tracted

I.

In a pair of recent papers, Karen Bennett has articulated and defended a solution to the causal exclusion problem (Bennett 2003, 2008). The exclusion problem (per Kim 1998, 2005) purports to show that non-reductive physicalists cannot vindicate mental causation. Bennett argues that the exclusion problem can be “tracted”, that mental and physical causes can be “compatible” with one another provided that they stand in a particularly intimate relation to one another.

Bennett’s compatibilist solution is similar to others that have been proposed, particularly that advanced by Stephen Yablo (1992) according to which the physical and mental are related as determinates are related to determinables. But Bennett rejects the determinate/determinables model, and instead argues for a different intimate relation between the mental and the physical. In this paper we argue that however attractive the compatibilist approach may be, Bennett’s proposal about the intimate relation between mental causes and their physical realizers cannot be correct.

II.

Bennett constructs the exclusion problem as arising from the mutual inconsistency of five independently plausible theses: distinctness, completeness, efficacy, nonoverdetermination, and exclusion (2008). For present purposes, the two theses we focus on are distinctness and exclusion (Bennett 2008: 281):

Distinctness: Mental properties (and perhaps events) are distinct from physical properties (or events).

Exclusion: No effect has more than one sufficient cause unless it is overdetermined.

As everyone knows—and as Kim (1998) urges—the causal exclusion problem can be evaded by identifying mental properties (or events) with physical properties (or events), or otherwise “reducing” the mental to the physical. That is to say, if we reject Distinctness then there is no need to worry about Exclusion, for the mental and physical do not even appear to be competing sufficient causes. Thus Exclusion is supposed to be a special problem for those who accept Distinctness—dualists and non-reductive physicalists.

Bennett’s compatibilist strategy, grossly put, is to argue that the mental and physical may be sufficiently tightly coupled that they fail to satisfy a plausible criteria for overdetermination, and without being identical.¹ In that case Distinctness will be maintained, but Exclusion is false: there is an effect that has more than one sufficient cause but is not overdetermined. The most familiar version of this strategy owes to Yablo (1992), who argues that if the mental and the physical are related as determinable and determinate then they do not overdetermine one another. The model for this explanation is the relation among colors, as between red and scarlet. Being red and being scarlet are plainly distinct properties, as it is possible to be red without being scarlet. But it is also plausible that when we stop at scarlet stop signs, the scarletness and the redness of the stop signs do not (normally) causally compete: one stops because the sign is red and also because it is a particular shade of red, viz., scarlet. The explanation for this is, very roughly, that being scarlet is a particular way of being red—or, perhaps better, that being red is

¹ Bennett (2003) argues that compatibilism is plausible. Bennett (2008) argues that the compatibilist solution is available to the non-reductive physicalist but not to the dualist.

part of being scarlet. (Anything that is scarlet is necessarily red. All members of the set of scarlet things are members of the set of red things. And so on.) So just as parts and wholes do not causally compete because they are distinct without being wholly distinct, so too determinates and their determinables do not causally compete because they are distinct but not wholly distinct.²

Bennett thinks that there are compelling reasons for denying that the mind-body relation is the determinable-determinate relation, and we agree (Bennett 2008; cf. Funkhouser 2006 and Haug 2010). She sets out to argue that there is a different intimate relation between the mental and the physical that can do the same work. Importantly, Bennett is not satisfied to merely assert the possibility of a compatibilist solution to the exclusion problem. Rather, she advances a specific proposal for how the physical-mental dependence relation that would vindicate compatibilism.

III.

Bennett's task is to show that the mental and the physical can each be sufficient causes of some effect without overdetermining that effect. The strategy is to argue that overdetermination only occurs when two counterfactuals are both nonvacuously true (Bennett 2008: 288):

(O1) if m had happened without p , e would still have happened: $(m \ \& \ \neg p) \ \Box \rightarrow$

e , and

(O2) if p had happened without m , e would still have happened: $(\neg m \ \& \ p) \ \Box \rightarrow e$.

If one or the other counterfactual is false or only vacuously true, Bennett argues, then there is no overdetermination. Crucially, (O2) would be only vacuously true if it were impossible for the physical cause to occur without the mental cause also occurring—as would be the case if, to give

² See Bernstein (forthcoming) for more on distinctness as it relates to overdetermination.

two examples, the mental and the physical were identical, or the physical were a determinate way of being a determinable mental state. But Bennett, we have seen, rejects those two options. Instead she argues that the presence of a sufficient physical cause necessitates the presence of the mental cause, or else the conditions that enable the physical property to have its effect also counterfactually guarantee that the physical property realizes the mental property. Consequently, (O2) is vacuously true or plainly false; so there is no causal overdetermination.

This result obtains, according to Bennett's proposal, because although the physical cause (p) of the effect (e) is distinct from e 's mental cause (m), "the conditions that must hold for p to bring about e —physical conditions, note—are *basically the same as the conditions in which p necessitates m* . So if p were to occur without m , those conditions would not hold—a p would not, or at least might not, cause e " (Bennett 2003: 488-489, italics original; see also Bennett 2008: 287-292). The idea is simple: On most views of causation, p is a sufficient cause of e only against some causal background conditions. This much is familiar. Similarly, if p realizes m , then it does so only under some appropriate conditions. This is less familiar, so will require some explaining, below. Bennett's insight is that if the conditions under which p causes e are "the same" as the conditions under which p realizes m , then (O2) will be at best vacuously true.

But what are the "conditions" under which p realizes m ? To use the language of Shoemaker (1981, 2007), if p is the "core realizer" of m , then p by itself is not metaphysically sufficient for m ; what is metaphysically sufficient for m is the "total realizer" of m which consists of the core realizer p plus some additional conditions that Shoemaker calls the "surround" (2007). For those who prefer to think of realization in terms of the satisfaction of Ramsey sentences, the point can be seen by remembering that the sufficient condition for realizing a Ramsey-Lewis functional state is *both* (a) being such that the Ramsified theory (abbreviated) T_x , holds *and* (b)

being in a physical state p that plays a specific role in Tx (Lewis 1970; Block 1978). It is those two properties together that are the total realizer of the state—that are metaphysically sufficient for the functional state. Merely having a core realizer p will not be sufficient for realizing the functional state unless it occurs in a system that also has the surround property, including being such that it satisfies the conditions Tx .³

The picture, then, is that p plus background conditions are causally (“horizontally”) sufficient for e , while p plus surround are metaphysically (“vertically”) sufficient for m . And Bennett’s idea is that if the background conditions are identical to or necessitate the surround, then the horizontal conditions are the same as or anyway sufficient for the vertical conditions. In that case, it will be impossible (at least in nearby possible worlds) for p to cause e without also realizing m . Consequently (O2) will either be false or only vacuously true. And Bennett supposes that it is clear that the background conditions are “the same” as the surround (Bennett 2008: 291). Therefore, she concludes, no causal overdetermination is occurring. Exclusion is tracted.

IV.

It is not the case that the causal background conditions are identical to the surround, and it is implausible that they necessitate the surround. The point can be made abstractly.⁴ But here we will demonstrate the incompatibility via a simple example. Following Ned Block (1978), let us think about realization and causation by considering the operations of vending machines of two different types, to be described below.

³ The surround may include (but not be identical to) the holding of Tx , because there may be further requirements presupposed by Tx but not included in it. (See Wilson 2001). Later that will be important; but for now it is only important that p alone is not the total realizer of m .

⁴ See AUTHOR 2010, forthcoming.

We take it that *being a vending machine* (of a specific sort) is a realized property *par excellence*. For simplicity, let us suppose that both devices that we will consider accept as payment only a unique token that is larger than all other coins or tokens in circulation.⁵ For machines of Type 1, the customer inserts a token and then presses the one available vending button, which causes a Coke to be dispensed into a bin at the bottom of the machine. For machines of Type 2, the customer inserts a token and then presses one of the two vending buttons, which cause a Coke or a Pepsi (respectively) to be dispensed into a bin at the bottom of the machine.

We shall consider two token devices, one of each type.⁶ Here is how Device 1 (of Type 1) operates. Device 1 realizes a Coke machine. When a token or coin is inserted, it slides down a slide and comes to a stop against a bumper. There is a hole in the bottom of the slide that has a circumference slightly smaller than the circumference of the proper token. If the inserted coin or token is smaller than the circumference of the hole, it falls through the hole and is returned to the customer via a return slot. If the inserted token is of sufficient circumference (if it is a token of the proper sort), it does not fall and instead remains at the bottom of the slide. At this point, the Coke machine is in the *Ready* state (alternatively, has the property of being *Ready*), and the vending button (Button 1) illuminates. When the vending button is pressed, an arm with a trigger on it swings in from the right toward the spot where a token would be if it has not fallen through a hole. If the coin is present, then the trigger is depressed (by contacting the solid surface of the token), and two things happen simultaneously: (a) the bumper moves out of the way and the

⁵ This idealization will allow us to describe especially simple physical realizers for the vending machines but is otherwise inessential.

⁶ We'll use "device" for the physical realizer properties and states, and "machine" for the realized properties and states.

token is dropped into the collection bin, and (b) a Coke is released from the storage area and drops (by gravity) into a retrieval bin at the bottom of the machine.⁷

Device 2 (of Type 2) realizes a Coke-or-Pepsi machine. Here is how Device 2 operates. Up until the point of the vending button pressing, Device 1 and Device 2 are identical. When the customer presses Button 1, the mechanisms in Device 2 operate exactly as in Device 1. But when Button 2 is pressed, an arm with a trigger on it slides in from the left (rather than right) toward the spot where the token would be if it has not fallen into a hole. If the proper token is present, then the trigger is depressed (again, by contact with the solid surface of the token), and two things happen simultaneously (a) the bumper moves out of the way and the token is dropped into the collection bin, and (b) a Pepsi is released from the storage area and drops (by gravity) into a retrieval bin at the bottom of the machine.

Let Q be the physical property of having a token resting at the bottom of the slide, which can occur in both Device 1 and Device 2. In Device 1 (realizer of machine Type 1) and Device 2 (realizer of machine Type 2), the occurrence of Q together with background conditions (especially saliently, the pressing of the Coke button) is causally sufficient for the dispensing of a Coke.⁸ Moreover, the background conditions for the dispensing of a Coke are the same for both machines: the machines must be plugged in to live outlets, there must be no blockage in the coin slot, there must be Cokes in the machines that are ready to dispense, the machines must be

⁷ Note that the inputs and outputs—tokens, Cokes and Pepsis, button presses—are not realized properties or states. They are “physical” properties relative to the realized machine properties. In the language of Lewis (1970), “Coke,” “Pepsi,” and “button press” are O-terms.

⁸ Here we use what Bennett calls a “permissive” (2003: 290) and “somewhat sloppy notion of causal sufficiency according to which reasonably normal physical events... count as causally sufficient” (2003: 489). Q itself is only sufficient for Coke dispensing given some other conditions, including the pressing of the vending button. But we can say, after Shoemaker (2007), that Q has the forward-looking conditional power of causing a Coke to be dispensed if Button 1 is pressed.

located in gravitational or inertial fields (so that the released beverage falls into the bin), and so on.

Furthermore, in both Device 1 and Device 2, the occurrence of Q is a core realizer of the vending machine state *Ready*. But having a token resting at the bottom of the slide is only a core realizer of *Ready* because it must occur in a suitably organized device in order to realize *Ready*. If the slide holding the token were removed and set on a table, there is nothing that would be *Ready*; and having Q in that way would not, combined with a button press, be sufficient for the dispensing of a Coke. Yet if the device parts are interchangeable, then the same slide and token could, at a different time, realize *Ready* in the same machine or in some other machine.

Although Q occurs in realizers of machines of both types, the total realizer of Ready (Q plus surround) is not the same for both machines. In machines of Type 2, Q must play a richer causal role such that, counterfactually, had the “Dispense Pepsi” button been pressed (realized by Button 2, in the above story), the result would be a Pepsi being dispensed. That is, Type 2 machines are such that had the second button been depressed, they would dispense a Pepsi. Being such that this Pepsi-dispense conditional holds is among the properties that must be in the surround in order for Type 2 machines to have total realizers for *Ready*. Type 2 machines that would not dispense a Pepsi were the Pepsi button to be pressed are not *Ready*.⁹ The Pepsi-dispense conditional must be true of Device 2 in order for the total theory of the machine (as described by the machine table) to be true of it, and thus in order for Q to realize *Ready* in that device.

⁹ Or else they are broken. We suppress the complication of defective realizers for simplicity.

One might worry about, for example, what happens if the machine is simply out of Pepsi but is has Coke, or vice versa. The answer is that the machine we described is not *Ready* in that case. But our machine is very simplistic and does not check for the presence of Pepsi, so you would lose your money if you pressed the Pepsi button and it was out of Pepsi.

In contrast, no such counterfactual properties need be part of the surround for *Ready* in machines of Type 1, because the dispensing of Pepsi is not among the outputs of Type 1 machines. This illustrates that although machines of Type 1 and Type 2 can have the same core realizer Q for the machine state *Ready*, the total realizers of *Ready* differ in Type 1 and Type 2 machines.

Our point can be easily seen by inspecting even simplified machine tables for Type 1 and Type 2 machines (Figure 1). In order for a device to realize the Ready state for a Type 1 machine, it must satisfy the complete machine table and, in addition, be in some physical state (e.g., Q) that has the relations definitive of *Ready* in the machine table, e.g., being caused by a token insertion and resulting in a Coke dispense if Button 1 is subsequently pressed. But in order for devices of Type 2 to realize *Ready*, all of that must be true and, moreover, they must be such that Q would cause a Pepsi to be dispensed were Button 2 to be pressed.

TYPE 1	Wait	Ready
Insert token	illuminate button; go to Ready	stay in Ready
Press Button 1	stay in Wait	dispense Coke; go to Wait

TYPE 2	Wait	Ready
Insert token	illuminate buttons; go to Ready	stay in Ready
Press Button 1	stay in Wait	dispense Coke; go to Wait
Press Button 2	stay in Wait	dispense Pepsi; go to Wait

Figure 1. Machine tables for Type 1 (Coke) and Type 2 (Coke or Pepsi) machines.

What the above shows is that the horizontal background conditions for the dispensing of a Coke are not the same as the vertical surround conditions for the realization of *Ready* in Type 2 machines. Thus surrounds are not identical to background conditions.

V.

Strictly speaking it is not true that *Ready* is a machine state for both Type 1 and Type 2 machines. It would be better to say that the two types of machines have analogous *Ready* states, *Ready-1* and *Ready-2*. Because the identity of a particular functional state is constituted by its total role, the whole machine table is relevant to the identity of every machine state. Thus no machine of Type 1 can ever be in *Ready-2*, and no machine of Type 2 can ever be in *Ready-1*.¹⁰

If this is right, then the following two conditions hold:

- (i) The causal (horizontal) background conditions for Q to be sufficient for effect *e* (dispensing a Coke) are the same in Device 1 and Device 2.
- (ii) The metaphysical (vertical) surround conditions for Q to realize *Ready-1* and *Ready-2* are different in Device 1 and Device 2.

It follows that the causal background conditions for the sufficiency of Q for *e* are not identical to the functional background conditions for the metaphysical sufficiency of Q for both *Ready-1* and *Ready-2*. In particular, Q together with Coke-dispensing causal-background conditions are not vertically sufficient for the realization of *Ready-2* even in Device 2.¹¹

¹⁰ It is worth emphasizing the holism inherent in realization theories. Functionalism, recall, is supposed to be an improvement over behaviorism because each functional state can be defined simultaneously with every other functional state, rather than circularly. As such, human “pain” can only be had by creatures with the entire psychological profile characteristic of humans, i.e. *Tx*. And the property named “Ready-1” can only be had by machines that fit the specific total functional profile of Type-1 machines.

¹¹ This will be the case even if the Type 2 machine is never actually in any of the states that differentiate it from a Type 1 machine.

VI.

We have argued that causal background conditions for cause p to have effects e are not invariably identical to the surround conditions for p to realize state m . We believe that surrounds and causal background conditions are never identical, a point to which we shall return. But first let us consider an alternative: Bennett (2003: 489) only claims that the background conditions and the surround are “basically the same,” not that they are identical. And her argument only requires that the causal background conditions are such that they, along with Q, necessitate machine states. So could it be the case that the causal background conditions necessitate the surround, even if they are not identical to it? No. Our two vending machines again illustrate why that cannot be the case.

In order for this weaker version of Bennett’s claim to be correct, it should turn out that if Q and the causal background conditions for the dispensing of Cokes necessitate *Ready-2* in machines of Type 2, then they also necessitate *Ready-2* in machines of Type 1. After all, if Q plus the relevant causal background conditions necessitate the total realizer of *Ready-2* in Type 2 machines, as Bennett requires, then they are metaphysically sufficient for *Ready-2*. Period. But that means that Device 1 (of Type 1) would realize *Ready-2*—for it has Q, and the relevant causal background conditions can hold for it—Q can cause the dispensing of a Coke. But it is impossible for Type 1 machines to have *Ready-2*—for dispensing Pepsi is not among their outputs, and all of the Pepsi-dispense counterfactuals are false of them. Yet satisfying those conditions is necessary for the realization of *Ready-2*. So Bennett is mistaken: the causal background conditions for Q to cause effect e (dispensing a Coke) do not necessitate the surround that is part (with Q) of the total realizer of *Ready-2*.

One might worry that, contrary to our claims, the counterfactuals distinctive of *Ready-2* are true of Device 1.¹² Because Device 1 has only one button, the thought goes, it is not *false* that it would dispense a Pepsi were Button 2 pressed, but rather *vacuously true*. So Q is sufficient for being in *Ready-2*, after all. We prefer to think of such Pepsi-dispensing counterfactuals as false for Device 1, which has no second button. But against this concern we can say several things. First, if the counterfactual about Button 2 is not false because the button is absent, then it is false because the nearest possible world in which Button 2 is pressed on Device 1 is a world in which the button is not connected to anything, so it is not true that a Pepsi would be dispensed—surely a world in which it has the button *and* the button is connected to the mechanism of the device is more distant. But, second, if one insists that the counterfactual is vacuously true of Device 1, then that is a reason to demand that the satisfiers of machine tables do so non-vacuously, or else we will be stuck with something close to computational universalism—a coffee cup vacuously satisfies the Machine 2 counterfactuals and thereby realizes *Ready-2*, but that is an unacceptable result. Finally, the appearance of a problem about these impossible antecedents may be an artifact of the presentation in terms of counterfactuals. If we think of the machine table as specifying laws rather than counterfactuals (as would be the case if we wrote the Ramsey sentences for the machine tables), then there is little temptation to think of the Machine 2 laws holding of Device 1.¹³ For all of these reasons, it is safe to conclude that Q plus the causal background conditions for dispensing a Coke are not, taken on their own, sufficient for the realization of either *Ready-1* or *Ready-2*.

¹² We are grateful to REMOVED FOR ANONYMOUS REVIEW for pushing us on this point.

¹³ There is nothing deep behind this thought, just the fact that counterfactuals with impossible antecedents can be true while laws with unsatisfiable antecedents are not.

VII.

We have argued that the causal background conditions for Q causing e are not always identical to and do not always necessitate the surround for Q being metaphysically sufficient for m . You might think that, even so, it could be the case that for some systems the background will be identical to or necessitate the surround. And if at least some cases of the physical realizing the mental are of those sorts, then the exclusion problem could still be tracted.

The first thing to say is that neither Bennett nor anyone else has given us any reason to think that the physical-mental realization cases are special in this respect. Indeed, one complaint about the exclusion problem is that it seems to generalize far beyond the mental (e.g., Bontly 2002). Likewise, Bennett's solution seems so attractive precisely because it does not involve any such special pleading on behalf of the mental. It would be, at the very least, surprising if the realizers of mental states happened to be conveniently distinctive in just the right way for compatibilism to work out.

The second thing to say in response is that there are good reasons to doubt that it could even sometimes be the case—we think it is not just surprising, but impossible. We don't have the space to set out the whole argument here, but we can give the gist.¹⁴ Start with causal background conditions. Background conditions for a sufficient cause are the kinds of factors that could themselves be causes—the presence of oxygen (for the striking of the match to cause the flame), being a woman (for the efficacy of birth control pills), and so on. In a currently popular vernacular, constituents of background conditions are “variables” that could be plotted in a causal graph except that they are not relevant to certain causal questions that concern us, perhaps because they are either very common (so taken for granted) or very rare (so infrequently

¹⁴ For the details, see AUTHOR (2010, forthcoming).

manipulated.) In short background conditions are among the causes of an effect *e*. In contrast, there is reason to doubt that surrounds are causal properties at all (McLaughlin 2009, AUTHOR 2010, Christensen and Kallestrup 2012). The reason is that surrounds are the conjunction of many properties including some that are not causal properties at all—for example, the property *being such that had different input been provided then different output would have been produced*.¹⁵ And it is plausible that conjunctive properties with non-causal properties as conjuncts are not themselves causal properties. But conjuncts such as these are exactly the sorts of properties that are the surrounds for core realizers. Background conditions for Q to cause *e* are, in contrast, just more of the kinds of variables that are cited in a causal explanation, perhaps suppressed because they are very common or very rare. Changing (“wiggling”) whether oxygen is present will change whether the match lights. But wiggling *being such that it would light at - 100°C* would not change whether the match lights. Yet, again, these are just the sorts of properties that go into the surround.¹⁶

Bennett herself notices that surrounds and background conditions are different kinds of properties:

There is, in short, an important mismatch between the sorts of physical properties and events that are typically invoked in instances of the exclusion problem, and those that constitute the supervenience base for the mental. It is only complicated

¹⁵ Earlier we explained the need for surrounds by reference to the Ramsified theory Tx that must hold in order for a system to realize any one of the functional states specified by the theory. But now we see, per footnote 3, that Tx is necessary but not sufficient for the total realizer, for that must also include such properties as the holding of the laws of nature (McLaughlin 2009).

¹⁶ Now the whole notion of a “causal property” as contrasted with “non-causal properties” is certainly in need of explication. (As a first pass, consider that two conjunctive properties that differed only in one of their conjuncts that was not a causal property could make no difference. So on counterfactual or difference-making accounts of causation or causal explanation, such properties would fail to be causes.) Suffice it for now to say that the burden is on the defender of Bennett’s strategy to show that causal background conditions are ever identical to or necessitate surrounds.

extrinsic physical properties, and physical events with complicated extrinsic essences, that will metaphysically necessitate mental ones. (2008: 290)

But Bennett draws the wrong conclusion. She focuses on the way that the need for extrinsic properties to be included in the total realizer (viz., that which metaphysically necessitates the mental property) explains how (O2) can be false—for p can occur without m if the surround is absent. But what she fails to see is that because of this “mismatch” the background conditions cannot include all the properties that are needed for the surround. It will not be true that, “The context within which the physical event or property guarantees the mental one *is the same as the background conditions within which it brings about its effects*” (2008: 291, italics original). It is not the case that the presence of the causal background conditions for a cause to have its effect will also be sufficient for the presence of the various conditions—including counterfactual conditions—that constitute the surround. Not only are they not identical, they are altogether different kinds of properties.

Although we have developed our critique of Bennett’s compatibilism in terms of machine functionalism, the point we are making is perfectly general. It applies to any theory that denies that p alone is metaphysically sufficient for m , but further requires that p stand in some additional “extrinsic” relations.¹⁷ Leaving aside machines, we can give general reasons for doubting that causal background conditions necessitate surrounds. First, causal background

¹⁷ If all those relations are actualized causal relations, then it will be p^* (p plus those relations) that is by itself sufficient for m . Then m is not reducible to p , but only because p was the wrong candidate— m is instead reducible to p^* . The sufficiency of p^* for m is compatible with Distinctness, and so not strictly speaking a form of reductionism as defined by Bennett. But if p^* by itself—without any surround—is metaphysically sufficient for m , then there is reason to doubt that p^* and m are wholly distinct. If property individuation is hyperintensional, then it is possible that some realized properties/states have modally unique realizers. Perhaps *being god* is the modally unique realizer of *being a perfect being*, or *being the number 2* is the modally unique realizer of *being the smallest prime number*. We don’t know of anyone who holds that the mental is not reducible to the physical but has the physical as its modally unique realizer. Whatever motivation one might have for such a view would be quite different than the usual motivations for non-reductive physicalists.

conditions are transitory while surrounds are not. Sydney Shoemaker (2007) explicitly states that core realizers are momentary properties while surrounds are “relatively stable properties” of the subject (2007: 21). This difference is, if anything, too modestly put. Consider the old example of c-fiber firing and pain. A person has c-fiber firing for a moment, typically for a period of time of the same duration as the pain itself. On the other hand, a person has the surround for pain for most of his or her life. The surround includes the property of satisfying the entire psychological theory in which pain is defined—in short, the property of being a psychological creature of the human sort.

In contrast to the relatively stable surround, the causal background conditions that fix the effects of the core realizer are themselves transitory. When a person puts their hand on a hot stove and her c-fibers fire, she may on one occasion “jump around swearing like a sailor” (Bennett 2008: 291), but she may on another occasion wave her hand about wildly—in polite company, say. On a third occasion she may simply grit her teeth or run to find ice to put on her fingers. In each case these are effects of the firing of the c-fibers. Why do the effects differ so markedly from occasion to occasion? Why do people not always react the same way to instantiations of the same core realizer? The answer involves, in part, differences in the causal background conditions. They are not stable, they are different on different occasions. Yet pain is realized each time, because the surround is stable.

Furthermore, surrounds are holistic while causal background conditions are not. The holism of surrounds is plain in our machine table examples, as well as in the work of David Lewis (e.g., 1970). Lewis advances an account of how theoretical terms get their content from the total content of the theories in which they are introduced. Thus if a theory introduces three new terms, each term gets its content from the whole theory, not merely from one-third of the

theory. It is for this reason that Lewis goes to great lengths to ensure that a theory can be true even if one of its components is not—or else all the terms so defined will be empty.¹⁸

Notationally, the holism and stability of the surround shows itself by the fact that the whole existentially quantified theory Tx appears in every theoretical definition derived from theory T . In the case of the psychological theory, the term ‘pain’ is defined as, say, the i -th member of an n -tuple such that the n -tuple satisfies Tx and the i -th role is the pain role.

We have established that the causal background conditions that fix the effects of the instantiation of a core realizer do not necessitate the presence of any surround. Note that our examples also show that there cannot be some third property, metaphysically prior to the surround and causal background, that necessitates both the surround and causal background.¹⁹ For if the surround were to obtain in virtue of such a prior necessitator that is also sufficient for the causal background conditions, then those causal background conditions would obtain whenever the subject has the surround. If that were the case, surround and background would not be modally independent, just as Bennett hopes. But we have demonstrated that the subject can have the surround in the absence of any particular set of causal background conditions—indeed, it is usual for the surround to be present without the causal background conditions for any given state also obtaining. It is not the case that the surround and background are counterfactually linked, so there is no common sufficient condition for surround and background.²⁰

¹⁸ Recall that Lewis ultimately appeals to brute force: taking the disjunction of the conjunctions of most of the laws (i.e. platitudes) in the theory (1972), to ensure that at least one disjunct is a conjunction of only true claims.

¹⁹ Our thanks to Carl Gillett for suggesting this possible saving maneuver.

²⁰ We focus on the case of the background necessitating the surround, because that is what is relevant to assessing (O2). You might wonder whether the surround could necessitate the background. If so, then one could try to argue that (O1) is only vacuously true. But Bennett, and advocates of Distinctness in general, take (O1) to be non-vacuously true on the grounds of multiple realization. Be that as it may, the surround does not necessitate the causal background conditions, and for the same reason that they do not have a common necessitator: it is common for the surround to be present without the causal background conditions.

VIII.

Bennett's strategy seems promising, and non-reductive physicalists ought to study her roadmap: physical property p and mental property m could be distinct without being counterfactually independent if there were some conditions that, when conjoined with p , both cause effect e and metaphysically necessitate m . We have shown that the causal background conditions for p to cause e are not the conditions that Bennett is looking for. The causal background conditions for some core realizer p (e.g., Q) to cause effect e (e.g., dispensing a Coke) do not necessitate the surround that is (with p) metaphysically sufficient for a realized state m (e.g., *Ready*). Causal background conditions are not identical to and do not necessitate surrounds. Consequently, Bennett has not shown how (O2) is always false or vacuously true for mental states and their physical realizers. Exclusion is not tracted.

We have not shown that there is no way for (O2) to be false or only vacuously true that is compatible with Distinctness. But, admirably, Bennett's goal is not merely to advance such a possibility but further to show how it could be the case. It could be the case if Yablo's determinate/determinable solution worked; but it does not. Bennett argues that the compatibilism is true, instead, because causal background conditions are "basically the same" as surrounds. She takes that connection to be a simple observation about background conditions and surrounds. But here we have demonstrated that she is not correct: surrounds are neither identical to nor necessitated by the background conditions for their core realizers to have causal effects.

This result is significant not only for Bennett but for any who hope to advance broadly compatibilist solutions to the exclusion problem. All of the non-reductive theories we know about appeal to relational properties such as those that constitute the surround. Indeed, one might

reasonably suppose that this relational feature is exactly what distinguishes non-reductive theories from their reductive cousins. But because those relations outstrip the local conditions for causal effects it is hard to see how the surround properties can be causally relevant to those effects, and thus hard to see how they could be guaranteed by local causal conditions. This is an old chestnut, not easily cracked.²¹

Of course there is a limiting case in which a core realizer p plus its causal background would be metaphysically sufficient for a realized state m : namely, the case in which p by itself is strictly sufficient for m . This would be so if $p = m$. What is less clear is whether there can be cases where $p \neq m$ and p by itself can necessitate m . If there is a compatibilist solution to the exclusion problem, we have not yet seen it.²²

References

AUTHOR

AUTHOR

Bennett, K. 2003. Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract It. *Noûs* 37 (3):471-97.

Bennett, K. 2008. Exclusion Again. In Jakob Hohwy & Jesper Kallestrup (eds.), *Being Reduced*. Oxford: Oxford University Press.

Berstein, S. forthcoming. Overdetermination Underdetermined.

Block, N. 1978. Troubles with functionalism. C. W. Savage (ed.), *Minnesota Studies in the Philosophy of Science, Vol. IX* (Minneapolis, MN: University of Minnesota Press)

²¹ See, e.g., Dretske (1988).

²² We are grateful for feedback from REMOVED FOR ANONYMOUS REVIEW. Authorship is equal, and authors names are listed alphabetically.

- Bontly, T. 2002. The Supervenience Argument Generalizes. *Philosophical Studies*, 109: 75-96.
- Christensen, J. and J. Kallestrup. 2012. Counterfactuals and Downward Causation: A Reply to Zhong. *Analysis*. [Online only, 7 June 2012]
- Dretske, F. 1988. *Explaining Behavior*. Cambridge, MA: The MIT Press.
- Funkhouser, E. 2006. The Determinable-Determinate Relation. *Noûs* 40: 548-569.
- Haug, M. 2010. Realization, Determination, and Mechanisms. *Philosophical Studies*. 150: 313-330.
- Kim, J. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.
- Kim, J. 2005. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.
- Lewis, D. 1970. How to Define Theoretical Terms. *Journal of Philosophy* 67 (13): 427-446.
- Lewis, D. 1972. Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, 50: 249–258.
- McLaughlin, B. 2009. Review: Sydney Shoemaker, *Physical Realization*. *Notre Dame Philosophical Reviews*. Retrieved from <http://ndpr.nd.edu/review.cfm?id=16607>
- Shoemaker, S. 1981. Some Varieties of Functionalism. *Philosophical Topics* 12, 1: 83-118.
- Shoemaker, S. 2007. *Physical Realization*. Oxford: Oxford University Press.
- Wilson, R. 2001. Two Views of Realization. *Philosophical Studies*, 104: 1-31.
- Yablo, S. 1992. Mental Causation. *Philosophical Review* 101: 245-80.