Excluding the Problem:
Bennett on Counterfactual Tests and Backtracking

Winner of the Gerritt and Edith Schipper Undergraduate Award for Outstanding Undergraduate Paper at the 61st Annual Meeting of the Florida Philosophical Association

Katelyn Hallman, University of North Florida

1. Introduction

The exclusion problem is a problem for any theory which holds that the mental is distinct from or irreducible to the physical.1 In brief, the problem is that if the mental and physical are distinct, and each is causally sufficient to bring about their effects, then our actions would frequently be overdetermined. However, since overdetermination doesn’t appear to be something that could happen frequently, the argument goes that one of the alleged causes of our actions (typically the mental cause) should be excluded.

In this paper, I begin by explaining and motivating Karen Bennett’s formulation of the exclusion problem. Bennett’s formulation of the problem is unique in that it’s not a pointed argument against any one particular view; rather, her formulation sets up the problem as a set of inconsistent claims, at least one of which must be denied to remove the inconsistency. I then explain and motivate Bennett’s solution. Bennett creates a counterfactual test for overdetermination, which is meant to show that the non-reductivist account of mental causation does not result in rampant overdetermination. Next, I explain and motivate a recent objection to Bennett’s solution by Chiwook Won. Won’s charge is that Bennett’s counterfactuals are not necessary for overdetermination. However, I will show that Won does not assess the counterfactuals in Bennett’s test correctly and, thus, Bennett’s solution remains viable.

2. Bennett’s Formulation of the Exclusion Problem

Throughout this paper, I follow Bennett’s presentation of the problem as a set of five individually plausible yet jointly inconsistent claims:
DISTINCTNESS: Mental properties (and perhaps events) are distinct from physical properties (or events).

Completeness: Every physical occurrence has a sufficient physical cause.

Efficacy: Mental events sometimes cause physical ones, and sometimes do so in virtue of their mental properties.

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

Non-overdetermination: The effects of mental causes are not systematically overdetermined; they are not on a par with the [standard cases of overdetermination].

DISTINCTNESS is the claim that the mental is neither numerically identical nor reducible to the physical. Support for DISTINCTNESS comes from thinking about qualia, which leads to the conclusion that even knowing all the physical facts about the universe would not tell us any facts about qualia due to its necessarily subjective and introspectively accessible nature.

Completeness is the claim that if an event is physical, it has a sufficient physical cause. It is not the claim that physical events only have physical causes (this is “causal closure of the physical”), nor that every event has a physical cause, nor does it require that one knows the complete causal story; rather, Completeness is only the claim that there are no gaps in the physical causal story. Completeness can be supported conceptually or empirically. The conceptual defense of Completeness comes from figuring out the correct definition of “physical.” According to this defense, the very meaning of ‘physical’ entails that if an event is ‘physical’ it must have a physical cause. Alternatively, empirical evidence seems to imply that there are no gaps in the physical causal story; thus we shouldn’t need to appeal to any other types of non-physical causes to explain physical events.

Efficacy is the claim that mental properties can (and sometimes do) actually cause physical events. Efficacy is supported by how we explain our behavior; we say things like “I went to the fridge because I thought there was soup in there.” We believe this way of explaining our behavior isn’t just a colloquial way of explaining what we observe—we genuinely think this is an accurate representation of reality. In other words, we believe that these explanations are literal. When I say “I laughed because I thought the joke was funny,” we mean that thinking the joke was funny literally caused the laughter; rather than the mental assessment being merely associated with the laughter.

Exclusion defines overdetermination (or ‘redundant causation’) as an event that has more than one distinct sufficient cause. Common examples of overdetermination include the following:

(i) Two people shoot at one person at the same exact time, the bullets both hitting
their mark at exactly same time, killing the person.\(^9\)

(ii) Lightning strikes a hay bale at the same time that a person throws a burning cigarette onto the same bale, causing it to catch fire. \(^10\)

(iii) Two people throw rocks at a window and each rock hits the window at the same exact time, shattering it. \(^11\)

In each of these examples, each cause was sufficient to guarantee the effect, but a second sufficient cause occurred, overdetermining the effect. \(^12\)

**Non-overdetermination** is the claim that while overdetermination could happen, it is not a frequent occurrence with the mental and physical. A metaphysical reason supporting *Non-overdetermination* is the idea that frequent overdetermination must be precluded by the correct theory of causation—whatever it may be. \(^13\) Another line of reasoning is that cases of overdetermination are fantastic coincidences, or are bizarre cases of “massive, unexplained correlation between the multiple causes;” therefore we cannot posit that they happen frequently. \(^14\) *Non-overdetermination* is also supported by the claim that frequently positing a second sufficient cause would be unjustified and 'gratuitous' as it goes against Ockham’s razor. \(^15\) Finally, *Non-overdetermination* is supported by the claim that even if overdetermination were to happen, it’s highly improbable that it would occur frequently. \(^16\)

Putting it all together, *Distinctness, Completeness, and Efficacy* have it that whenever a physical event, e, was caused by a mental event, \(\varepsilon\) must also have a distinct physical cause. Adding *Exclusion*, if it were the case that physical causes and mental causes were both separately sufficient for the event, then overdetermination would happen *extremely* often. But, according to *Non-overdetermination*, overdetermination cannot happen frequently. Thus, we have an inconsistent set of claims. Something has to give, but figuring out which claim to deny will not be easy and will not come without a cost.

3. Bennett’s Solution

In her paper “Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It,” Bennett defends her version of a view that Terence Horgan calls “causal compatibilism” (just “compatibilism” from here, on). \(^17\) All supporters of compatibilism agree with *Distinctness, Completeness, Efficacy* and *Non-overdetermination*, but deny *Exclusion*. The denial of *Exclusion* amounts to the claim that the mental and physical can be distinct, sufficient causes without resulting in overdetermination. To support her view, the compatibilist needs to show that, while the ‘textbook examples’ (e.g. firing squads) would count as overdetermination, normal situations involving mental and physical causation do not count as overdetermination. According to
Bennett, “The compatibilist needs to break the analogy between the two types of cases.”

3.1.1. Bennett’s Counterfactual Test for Overdetermination

Bennett’s solution involves providing a counterfactual test detailing a necessary condition for overdetermination and using this test to show that cases of mental causation do not meet the counterfactual conditions. Recall from EXCLUSION that overdetermination occurs when an event has two distinct and sufficient causes. With this in mind, Bennett creates the following counterfactual test:

\[ c_1 \land c_2 \text{ overdetermine some event } \varepsilon \text{ only if:} \]
1. If \( c_1 \) happened without \( c_2 \), \( \varepsilon \) still would have happened, and
2. If \( c_2 \) happened without \( c_1 \), \( \varepsilon \) still would have happened.

This test only provides a necessary condition; thus, it can only be used to figure out which cases of causation do not qualify as overdetermined. Since this test is not sufficient for overdetermination, some events may fit the requirements and yet not be overdetermined.

It is important to note that this particular counterfactual test should be assessed in a non-typical way; counterfactuals are typically assessed by removing a part of the conditional and replacing it with something different (i.e. backtracking), whereas Bennett’s counterfactuals are to be assessed by deleting part of the conditional (one cause) and not replacing it with anything else (not another cause, or a cause similar to the deleted one). Consider the following example, “if I didn’t say that one thing, I would have nailed that interview.” There are two ways we can think of this counterfactual. The first way of thinking about this counterfactual would look like this: “In a world where I didn’t say that one thing, I would have awkwardly rambled, so I would have ruined that interview too;” this way of looking at the counterfactual involves backtracking. But, the second way of looking at this counterfactual involves keeping this world exactly the same and entirely deleting the fact that I said that one thing—then determining whether I would have succeeded in the interview. This way of looking at counterfactuals is how Bennett thinks we should assess the counterfactuals in her test. She argues that, in some contexts, backtracking is “definitely inappropriate”—and her test is one of those situations.

Since the goal of her test is to determine whether mental causation results in frequent overdetermination, Bennett considers her counterfactuals in terms of the mental and physical. Bennett replaces \( c_1 \) with “\( m \)” for “mental cause,” and \( c_2 \) with “\( p \)” for “physical cause.” Her test, using these variables, is as follows:
m & p overdetermine some event ε only if:

(O1) If m happened without p, ε still would have happened, and

(O2) If p happened without m, ε still would have happened. 23

Using this test, Bennett aims to deny EXCLUSION by showing that cases of mental causation do not result in frequent overdetermination. 24 To show this, Bennett argues we must show that at least one of the two counterfactuals is either false or vacuously true.

3.1.2. Showing the Falsity of the Counterfactuals

Bennett first considers denying that mental causation is overdetermination by showing that at least one of the counterfactuals is false. However, she argues that this attempt is not favorable because the falsity of these counterfactuals undermines EFFICACY and COMPLETENESS.

The falsity of (O1) entails that mental events need to be coupled with physical events to cause a physical event. This is not optimal, Bennett argues, because it undermines EFFICACY. 25 By denying EFFICACY, we would have to dramatically change the way we explain behavior. We explain behavior in terms of the mental, sometimes, because we actually think these explanations are correct and importantly different from saying “the sun rises.” Saying “the sun rises” is simply a colloquial way of explaining what we observe; but when we say that mental properties are causal properties, we think this is literally true and an accurate depiction of reality.

The falsity of (O2) entails that the physical needs help from the mental to be causally efficacious. This, though, seems to go against COMPLETENESS. If we were to deny COMPLETENESS, we’d appear to stand opposed to the scientific project of seeking out physical causes and explanations. What scientists strive to do is to figure out what the cause is not whether there was a cause.

3.1.3. Showing the Vacuity of the Counterfactuals: Bennett’s Preferred Solution

Bennett then considers denying that mental causation is overdetermination by showing that at least one of the counterfactuals is vacuously true. 26 Bennett’s final solution ultimately sides with arguing for the vacuity of (O2).

For (O1) to be vacuous, one needs to argue that it is impossible for m to occur without p. This, however, undermines multiple realizability, which is the idea that the same kind of mental state can be realized by different physical states. 27 By arguing for the vacuity of (O1), we are claiming that a certain mental state can only happen when a creature is in one exact physical state and that any
creature that is not in this exact physical state is not experiencing the mental state in question. Without mental states being multiply realizable, we cannot claim that different people or species have the same mental states as us. For example, if our mental states are not multiply realized in other species we could not say that our dogs feel sad when we leave home or that our cats enjoy it when you pet them because their brains have different structures than ours.28

For (O2) to be vacuous, one needs only to argue that it is impossible for p to happen without m also occurring. This, according to Bennett, seems more reasonable than the previous options because it coincides with the typical non-reductivist claim made in regards to the mind-body relationship: that there is an “upwards necessitation relation” between the physical and mental. According to Bennett, physicalists of all varieties must argue that once the physical facts are set, all the facts about mental states are set and that it is metaphysically necessary for everything to globally supervene on the physical. According to supervenience, in any possible world where A supervenes on B, it is impossible for there to be a world with B and not A. Supervenience requires that no two beings could be exactly alike physically while being different mentally.29 This view is exactly what claiming the vacuity of (O2) entails: that there is no possible world where p occurs and m does not occur. Given this, Bennett argues that all physicalists must claim that (O2) is vacuously true.30 By claiming the vacuity of (O2), Bennett shows that mental and physical causes can still be distinct and sufficient causes without resulting in overdetermination. Mental causation does not result in frequent overdetermination because the mental and physical are not independent of each other due to the necessary supervenience relationship—they are distinct, but necessarily linked.31 Using this reasoning, Bennett establishes a difference between firing squad overdetermination and mental causation; where the firing squad members are independent causes, mental and physical causes are not. According to Bennett’s solution, EXCLUSION is false, normal cases of mental causation are not cases of overdetermination, which resolves the inconsistency in the exclusion problem.32

4. Objection: Bennett’s Test Is Not Necessary for Overdetermination

Chiwook Won, in “Overdetermination, Counterfactuals, and Mental Causation,” argues that Bennett’s test is not necessary for overdetermination. Thus, Won’s charge is that her solution relies on a false premise.33

Won presents cases that he claims are cases of overdetermination, yet they do not satisfy Bennett’s counterfactuals. Won uses an example similar to the standard preemption examples:34

Bottle Shattering:
Two children, Billy and Sally, throw rocks at a glass bottle. Unbeknownst to Billy and Sally, there’s a bystander: Charlie. Charlie knows that Billy has a sensitive ego and that if Sally’s
rock breaks the bottle and Billy misses, he will be really upset. So, Charlie plans that if Billy’s aim is not accurate, he will throw a rock to stop Sally’s rock from hitting the bottle. But, Billy’s aim is true and both of their rocks hit and shatter the bottle.35

Won argues that Billy’s and Sally’s rocks overdetermine the bottle shattering because both of their rock-throwing events are distinct and sufficient causes for the bottle shattering. But, this situation does not satisfy Bennett’s counterfactuals because of Charlie’s inactive presence. It is true that if Billy’s rock hit the glass bottle and Sally’s did not, the glass bottle would have shattered. However, because of Charlie, if Sally threw the rock with good aim and Billy didn’t, Charlie would have intervened with Sally’s rock and the bottle would not have shattered. Charlie’s intended interference precluded the truth of one of the counterfactuals. Thus, with Bottle Shattering, it looks like Bennett’s test doesn’t properly diagnose overdetermination.

Won argues that there are also cases of overdetermination that involve additional trumping causes that don’t directly interfere with either of the overdetermining causes.36 To show this, he slightly modifies Schaffer’s trumping example:

Major Trump:
Two sergeants give orders to a squad of soldiers. Both the sergeants shout “Advance” at the same time, and the soldiers advance. Now suppose that a major was standing there too. He was actually about to order a retreat. But the major hears the two sergeants ordering an advance, so he does nothing. However, if only one of the sergeants had ordered the advance, he would have ordered a retreat, which would have caused the soldiers to retreat.37

In this case, according to Won, the advancing of the soldiers is overdetermined but neither of Bennett’s counterfactual is true because, if only one sergeant ordered an advance, the major would have ordered the soldiers to retreat. The major, unlike the case involving prevention above, does not directly interfere and cut off the sergeants’ orders to advance; rather, if one of the sergeants didn’t order, the major would trump the other’s order. Bennett’s test is supposed to show that situations that don’t satisfy her test are not overdetermination, but Won purports to show that there are cases of genuine overdetermination that do not satisfy the conditions of her test. If Won’s counterexamples successfully show what he argues, then Bennett’s test does not show that mental causation isn’t overdetermination.

5. Response: Backtracking and Independence

Won’s argument involves adding third parties into the causal stories that, while causally inactive,
prevent the truth of one or both of the counterfactuals. While this would be a good strategy if these counterfactuals were able to fall prey to backtracking, Bennett explicitly states that backtracking is not the appropriate way to think of her counterfactuals in the following quote:

Now, backtracking evaluations are not always and everywhere wrong, but they are definitely inappropriate in some contexts, and I hereby claim that this is one of them. To get the proper results from the overdetermination test, you cannot backtrack, looking for the reason the one event failed to occur. You can just imagine its failure to occur, period. So even in this case [where the first gunman didn’t shoot causing the second gunman to miss], if the first gunman had not fired but the second had, the victim would still have died. This case does not constitute a counterexample to the necessity claim.38

Bennett argues that (in terms of the mental, m, and physical, p), when thinking about m happening without p, we shouldn’t think of this as meaning removing p and replacing it with something similar to p (p*), nor should we think about the nearest worlds where p doesn’t take place, nor should we add any additional reasons or causes to explain why p doesn’t occur. Rather, we should think of this as deleting p from the story and leaving it empty.39 Bennett’s counterfactuals simply should not be analyzed using backtracking.

Won seems to indirectly address Bennett’s claim above, but he thinks that such a move is ad hoc. Bennett doesn’t give reasons to support why one shouldn’t backtrack when thinking about her counterfactuals; she just “hereby claim[s]” that one shouldn’t—and moves on.40 This may make it seem like she’s only making the claim to protect her test from backtracking counterexamples, which is ad hoc. The onus should be on the compatibilist (Bennett) to argue and explain why backtracking isn’t appropriate when thinking about her counterfactuals. Bennett’s whole reason for writing her initial paper was based on the claim that compatibilists must argue and explain why the view is true. However, it may be that she makes the same mistake when she doesn’t adequately support her claim that backtracking would be inappropriate for her counterfactuals.

However, Bennett’s claim is not ad hoc. Since Bennett doesn’t explicitly note why one shouldn’t backtrack, I provide a novel and helpful way to understand Bennett’s counterfactuals. One should not backtrack when assessing Bennett’s counterfactual test because her test is best thought of as testing the independence of the causes. Her test is meant to show that an event can have two distinct yet sufficient causes without resulting in overdetermination; this happens when distinct and sufficient causes are necessarily non-independent causes.41

By appealing to the necessary supervenience relationship between the mental and physical, Bennett shows that the mental is necessarily non-independent from the physical. Because it is impossible for the mental to exist without the physical, the mental does not exist independent of the
physical. Since *independence* is being tested in Bennett’s counterfactuals, adding or replacing causes muddles the story, which makes it impossible to show how c1 was *independent* from c2. We can make an analogy between this and how we do science. When conducting scientific experiments, researchers try to limit for all (or as many as possible) confounding variables, or variables that could interfere with the object of study. For example, if a psychobiologist wanted to study whether testosterone, independent of serotonin, influenced aggression, the researcher would manipulate *only* the levels of testosterone and hold the serotonin levels constant—they would *not* manipulate both variables. If you manipulated both variables, you would then be unable to conclude whether testosterone, individually, was the source of the change (or lack thereof) in aggression. When determining whether two things are able to be causes independent of each other, only one cause can be manipulated or changed while the other cause and variables must remain constant. Similarly, backtracking (manipulating and adding extra variables) should not be used when assessing Bennett’s counterfactuals.

Bennett’s emphasis on supervenience and *non-independence* breaks the analogy between mental causation and the firing squad. The mental and the physical have a necessary supervenience relationship that makes them *non-independent* of each other; whereas the individuals in the firing squad are not necessarily linked and are thus *independent* causes. While the individuals in the firing squad could have a relationship such that, if one didn’t fire, then the other didn’t fire, or perhaps if they are related in that they will only shoot if ordered to, this is only a contingent relationship. Bennett’s test shows that only cases where both causes are *necessarily* dependent on each other will not result in overdetermination.

6. Concluding Bennett’s Solution

Bennett’s solution to the exclusion problem sets out to deny *Exclusion* by showing that there are cases where an event is caused by two distinct and sufficient causes without resulting in overdetermination. To show this, Bennett uses supervenience to show that the mental and physical are necessarily not *independent* of each other. This necessary link, which bars *independence*, is what keeps mental causes from being cases of inherent overdetermination, systematic or not.

The upshot of Bennett’s argument is that distinct and sufficient causes don’t result in overdetermination when they’re not *independent* causes. Thus, a revised and more appropriate definition of *Exclusion* would be:

*Exclusion*: No effect has more than one *independent* sufficient cause unless it is overdetermined.
This formulation of EXCLUSION would allow mental causation to avoid the exclusion problem while still maintaining the integrity of the textbook cases of overdetermination. Bennett’s counterfactual test sets out to show that, while mental and physical causes are distinct and sufficient, the necessary link barring independence between the mental and physical results in the vacuity of (O2). And, since her test is necessary for overdetermination, the vacuity of (O2) shows how mental causation does not result in overdetermination.

Won’s criticism, because it relies on backtracking, does not properly assess Bennett’s counterfactuals. Bennett’s counterfactual test aims to demonstrate the independence of causes, and backtracking is not appropriate for testing whether something is independently sufficient for an event. Therefore, Bennett’s solution to the exclusion problem is not threatened by Won’s criticisms and remains a viable solution to the exclusion problem.

Notes

1 The problem is most commonly brought up against property dualism and non-reductive physicalism.
2 K. Bennet., “Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it,” Noûs 37.3 (2003): 471-472. For ease of explication, I will discuss the problem in terms of mental/physical states, properties, and events. Everything I say about the problem, however, can also be applied to mental/physical substances.
10 Mackie, The Cement of the Universe, 44.
Note that EXCLUSION does not make a judgment on how often overdetermination does or should happen.


14 Sider, “What’s so bad about overdetermination?” 722-723.

15 Sider, “What’s so bad about overdetermination?” 724.


18 Bennett, “Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it,” 474.


20 Bennett, “Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it,” 480.

21 Bennett, “Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it,” 482, 483, 484, and 488.

22 Bennett, “Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it,” 478.

23 Bennett, “Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it,” 480.

24 Bennett, “Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it,” 480.

25 Bennett, “Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it,” 481.

26 All conditionals that start with a false antecedent will be true, regardless of their consequent—and these are the conditionals that are vacuous. While they may be true, their truth is negligible because of the false antecedent.


28 Another potential drawback of claiming the vacuity of (O1) is that this denial entails that physicalism is necessarily true. But, Bennett argues that physicalism is plausibly only contingently true; it seems at least possible for physicalism to be false in other possible worlds (see Bennett, (2003) 483-484; Lewis, (1983) 362; Chalmers, (1996) 41-42; Jackson, (1968) 11-12). Additionally, the necessity of physicalism means that there can be no universe where there are mental states alone, which implies that it is impossible for immaterial souls to exist.

30 Bennett, “Mental Causation,” 4-10
31 Bennett, “Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it,” 476-477.
32 Does Bennett’s solution count as a denial of DISTINCTNESS? It does not; two things can be distinct and still dependent on each other. Consider the equation E = mc²; this defines how we calculate the kinetic energy of an object (by multiplying the object's mass by the speed of light squared). According to this equation, the kinetic energy that an object has is dependent on the object's mass and the speed of light; if the mass of the object were different or the speed of light were different, an object’s kinetic energy would be different. While kinetic energy depends on m and c, an object’s kinetic energy is not identical to the objects mass and the speed of light (it is distinct from these things). Thus, it appears that there can be dependence relation two things and the things still be distinct.
36 For more on trumping, see: Schaffer, (2000) 165; Menzies (2014).
39 Bennett, “Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it,” 478.
40 Bennett, “Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it,” 478.
41 For the rest of the discussion on Bennett, we should understand ‘independence’ as not necessarily bound together. The “necessary” part is important because, at best, Bennett only shows that cases of mental causation, where the mental and physical are necessarily non-independent causes, is not overdetermination. I am using independence to denote things that are necessarily non-independent due to some necessary relationship. However, the mere fact that there are some cases where two causes are contingently independent may not be enough to save these cases from being overdetermination. See: Chiwook Won, “Overdetermination, Counterfactuals, and Mental Causation.”

Bibliography


Panchakunathorn, P, “Overdetermination and the Exclusion Problem in Mental Causation”, academia.edu, accessed March 01, 2015:
http://www.academia.edu/2950535/Overdetermination_and_the_Exclusion_Problem_in_mental_causation


