Causal Closure and the Exclusion Argument
Gabriel Greenberg (UCLA), February 2, 2019

The Exclusion Argument is one of the main arguments for physicalism, and against interactionist dualism. In this note, I’ll explain the argument, provide some background about causation, and elaborate on the key premises.

1 Causation

The Exclusion Argument is all about causation, so we’ll first have to get clear about a few background assumptions having to do with the structure of causation.

To begin, it is widely assumed that all physical events (except perhaps for the beginning of the universe) have causes. In fact, nearly all physical events have many causes. Take the recent wildfires in Los Angeles. Why did this event occur? One cause was (let’s imagine) a sparking power line. Another cause was the dryness of the wood in the area, itself caused by prolonged drought. Perhaps a third cause was the lack of rain during the time of the fire. And so on.

Let us call these different factors PARTIAL CAUSES. They are causes of the wildfire. But they are only partial, because none of them alone would have been enough to bring about the fire. For example, if the forest had been wet, the sparking power line would not have been enough to cause the fire.

Although each partial cause alone is insufficient to lead to fire, all the partial causes together are sufficient. In fact, we can collect these many partial causes together into a giant bundle of events. I’ll call this bundle of events a FULL CAUSE. Let us define the full cause of an event as any collection of partial causes which together are sufficient to bring about that event. Some authors use the term “sufficient cause” for the same idea.

While most large-scale events have many partial causes (like the wildfire), it is rare for an event to have more than one full cause. When an event has more than one full cause, we call it a case of OVERDETERMINATION. Overdetermination occurs only in unusual or coincidental situations. Take the case of two assassins who both shoot the same target in the heart, at the same time, but from different positions. Both shots would have killed the target on their own, so both shots are full causes. Yet clearly this is abnormal case; most events are not overdetermined.

The point for now, is not to confuse an event having multiple partial causes, which is normal, with an event having multiple full causes, which is highly unusual.
2 Causal closure

The causal closure of physics is the idea that every physical event has a full (or sufficient) cause, and that this cause is itself physical. Crane (2001, 45) puts it this way: “Every physical event has a physical cause which is enough to bring it about, given the laws of physics.” The idea is captured well by Papineau (not part of the original reading):

At first pass the causal closure of physics says that every physical effect has a sufficient physical cause... If you consider any physical effect, then there will arguably always be some prior sufficient physical cause: we expect to be able to account for physical effects without leaving the physical realm itself. In particular, this seems to hold even for physical effects which take place within the bodies of conscious beings. When the muscle fibres in my arm contract, this is presumably due to electrochemical activity in my nerves, which is due to prior physical activity in my motor cortex, and so on. In principle, it would seem possible to account for this entire sequence solely in terms of the resources offered by physics itself, and without making any essential appeal to any other subject matter. (Papineau 2009, 54)

This is the same idea that Carruthers is getting at in this passage:

Most scientists now believe that physics is closed, in the sense of permitting no interference from, or causation by, events at higher levels of description (e.g., chemical or biological). On this view, all atomic and sub-atomic events happen in accordance with physical laws (albeit probabilistic ones), and all events at higher, more abstract, levels of description must be realized in, or constituted by, those physical processes, in such a way as to allow no independent point of causal leverage. (Carruthers 2004, 150)

Papineau goes on to explain the significance of causal closure for philosophy of mind:

The causal closure of physics is solely a claim about how things go within physics itself. It does not assert that everything is physical, but only that those things that are physical have a physical cause. So it does not rule out realms of reality that are quite distinct from the physical realm. It is entirely consistent with the causal closure of physics that there should be self-sufficient realms that operate quite independently of physical goings-on (a realm of ghostly spirits, say). The causal closure of the physical says only that when we are within the physical realm we will find that every physical effect has a physical cause.

Even so, the causal closure of the physical does give rise to a powerful argument for reducing many prima facie non-physical realms to physics: for it indicates that anything that has a causal impact on the physical realm must itself be physical. This is because the causal closure of the physical seems to leave no room for anything non-physical to make a causal difference to the physical realm, since it specifies that every physical effect already has a physical cause. (Papineau 2009, 54)
He concludes:

The thesis of the causal closure of the physical thus argues that many prima facie non-physical occurrences—all those that exert an influence on the physical realm—must themselves in fact be physical. For otherwise it is hard to see how they could have any physical effects. (Papineau 2009, 55)

There are various approaches to defending causal closure, each taking off from some aspect of contemporary scientific or technological practice. Carruthers emphasizes the general explanatory success of the physical sciences:

What reason do we have for believing in the causal closure of physics? This is not something which can be proved (least of all by thought alone, of course). But for some centuries it has been a successful methodological assumption of scientific enquiry. Scientists work under the assumption that processes in physics brook no interference from higher levels of causation. Whenever they come across physical phenomena which cannot presently be explained in physical terms, instead of postulating causation by Œlan vital (a supposed independent biological life-force), or causation by ectoplasm (a supposed independent psychic force), or whatever, they look deeper into the physical mechanisms. In many such cases this deeper look has proved successful; and in all such cases physicalistic scientific enquiries continue to make progress. This gives us good reason to think that the scientific methodology is correct, and that physics is indeed closed. (Carruthers 2004, 150)

Papineau emphasizes growing confirmation from brain science:

The most significant evidence seems to have come directly from physiology and molecular biology, rather than from physics. Over the last hundred and fifty years a great deal has come to be known about the workings of biological systems (including brains), and there has been no indication that anything other than basic physical forces is needed to account for their operation. In particular, the twentieth century has seen an explosion of knowledge about processes occurring within cells, and here too there is no evidence of anything other than familiar physical chemistry. The result has been that the overwhelming majority of scientists now reject vital and mental forces, and accept the causal closure of the physical realm. (Papineau 2009, 57)

3 The Exclusion Argument

The Exclusion Argument has three premises:

1. **Downward Causation:** Many mental events cause physical events.

2. **Causal Closure:** All physical events (that have causes) are fully caused by sequences of prior physical events.

§3 The Exclusion Argument
3. **Exclusion:** No event is fully caused by more than one independent sequence of prior events (except for rare case of causal coincidence).

Collectively, these premises imply that dualism is false. How do we get to this conclusion?

First, Downward Causation tell us that there will be mental events that cause physical events. How should we interpret this? Is the causation in question *partial* or *full*? Mental events are always going to combine with background conditions to cause physical events. (Think again of the wildfire.) So it’s probably best to think of Downward Causation as saying that mental events are essential parts of the full causes of physical events.\(^1\) Carruthers goes exactly this direction in his statement of Downward Causation: “Our decisions sometimes form part of the true causal explanation of some of our movements.” (Carruthers 2004, 158)

For the sake of exposition, let’s focus on a particular case. In a given case of downward causation, let’s call the mental event \(M\) and the physical event \(P\).

Next, Causal Closure tell us that for every physical event, there is another sequence of events which (i) is a full cause of the first, and (ii) is physical. Now we can relate this observation to Downward Causation. It implies that, given the physical event \(P\), there is another sequence of events \(P^*\) such that (i) \(P^*\) is a full cause of \(P\), and (ii) \(P^*\) is physical.

At this point, we’ve derived that \(P\) seems to have two full causes: one that includes \(M\), on one hand, and \(P^*\) on the other. But Exclusion tells us that we are headed for trouble. According to Exclusion, no event has two full causes (with some exceptions—I’ll return to that in a minute.). This seems to imply that \(P^*\) and the sequence including \(M\) can’t both be full causes of \(P\). But does it? Actually all that Exclusion says is that there can’t be *more than one* full cause of \(P\). That requirement is compatible with the first two premises, but only if the sequence including \(M\) and \(P^*\) are the same. And that is exactly what the defender of the Exclusion argument says: \(M = \text{part of } P^*\). And more generally, in every case of downward causation, the mental event involved is *the same as* one of the physical process which led up to the final event.\(^2\)

Here is the way Carruthers points the point:

> Now the only way in which we can hold onto both beliefs — the belief that some mental events are causally necessary for the occurrence of some physical ones, and the belief that it is unnecessary to appeal to anything other than physical events in providing causal explanations of brain events — is by believing that some mental events are physical ones. (Carruthers 2004, 153)

If every case of downward causation involves mental events which are also physical, then dualism can’t be true. This is simply because dualism claims that all mental events are non-physical.

---

\(^1\)Writing tip: in a short essay, it’s probably easier to treat Downward Causation as saying that mental events are full causes of physical events. This will make the rest of the argument go a little more directly, and it’s a harmless simplification.

\(^2\)Note: in class I presented a slightly simplified version of this argument. If we assume that Downward Causation says that mental events are *full* causes of the physical events, then we can conclude that \(M = P^*\), rather than \(M = \text{a part of } P^*\). Either presentation is fine for the purposes of your papers.
From here, we can vault ourselves to the conclusion that physicalism itself is true, with the help of an additional assumption. The assumption is that all mental states have the same basic metaphysical nature. This idea is certainly debatable, but I’m going to let it stand for now. By way of motivation, it’s worth nothing that, for the dualist, if anything is non-physical, its the decisions that figure in downward causation. So if decisions turn out to be physical, even the interactionist dualist is likely to concede the rest of the mind to the physicalist. (Actually, consciousness introduces a wrinkle into this story; we’ll come back to this point at the end of the course.)

4 Papineau on the Exclusion Argument

I want to quote in full Papineau’s (2004, 17-18) explanation of the Exclusion Argument, which wasn’t part of the original reading. It covers the same ground as my account above, but I think it’s helpful to hear the same argument from different points of view. (For yet another statement of the argument, see Crane 2001, pg. 49-50.)

Let me now outline what I take to be the canonical argument for materialism. Setting to one side all complications, which can be discussed later, it can be put as follows.

Many effects that we attribute to conscious causes have full physical causes.
But it would be absurd to suppose that these effects are caused twice over.
So the conscious causes must be identical to some part of those physical causes.

To appreciate the force of this argument, consider some bodily behaviour which we would standardly attribute to conscious causes. For example, I walk to the fridge to get a beer, because I consciously feel thirsty. Now combine this example with the thought that, according to modern physical science, such bodily movements are fully caused by prior physical processes in brains and nerves. The obvious conclusion is that the conscious thirst must be identical with some part of those physical processes.

Let me now lay out the above argument more formally. This will help us to appreciate both its strengths and its weaknesses.

As a first premiss, take:

(1) Conscious mental occurrences have physical effects.
As I said, the most obvious examples are cases where our conscious feelings and other mental states cause our behaviour.
Now add in this premiss (‘the completeness of physics’ henceforth):

(2) All physical effects are fully caused by purely physical prior histories.
In particular, this covers the behavioural effects of conscious causes to which our attention is drawn by premiss 1. The thought behind premiss 2 is that such physical
behaviour will always be fully caused by physical contractions in your muscles, in turn
caused by electrical messages travelling down your nerves, themselves due to physical
activity in your motor cortex, in turn caused by physical activity in your sensory cortex,
and so on.

At first sight, premisses 1 and 2 seem to suggest that a certain range of physical ef-
effects (physical behaviour) will have two distinct causes: one involving a conscious state
(your thirst, say), and the other consisting of purely physical states (neuronal firings,
say).

Now, some events are indeed overdetermined in this way, like the death of a man who
is simultaneously shot and struck by lightning. But this seems the wrong model for
mental causation. After all, overdetermination implies that even if one cause had been
absent, the result would still have occurred because of the other cause (the man would
still have died even if he hadn’t been shot, or, alternatively, even if he hadn’t been struck
by lightning). But it seems wrong to say that I would still have walked to the fridge even
if I hadn’t felt thirsty (because my neurons were firing), or, alternatively, that I would
still have gone to the fridge even if my neurons hadn’t been firing (because I felt thirsty).
So let us add the further premiss:

(3) The physical effects of conscious causes aren’t always overdetermined by
distinct causes.

Materialism now follows. Premisses 1 and 2 tell us that certain effects have a conscious
cause and a physical cause. Premiss 3 tells us that they don’t have two distinct causes.
The only possibility left is that the conscious occurrences mentioned in (1) must be iden-
tical with some part of the physical causes mentioned in (2). This respects both (1) and
(2), yet avoids the implication of overdetermination, since (1) and (2) no longer imply
*distinct* causes.

## 5 Understanding Exclusion

Now, I want to take a slightly closer look at the Exclusion premise, which admittedly can be confus-
ing. The way I stated it was: “No event is fully caused by more than one independent sequence of
prior events (except for rare case of causal coincidence).” (Papineau’s premise (3) above expresses
the same idea, with a narrower focus.)

First, I want to stress that Exclusion is *not* about *partial* causation. As I discussed above, most events
have many partial causes, and Exclusion has no problem with that. Instead, Exclusion is only about
*full* causation. It states that, normally, events have only full cause.

Second, despite the fact that Exclusion has an abstract and strange sound to it, I think it is actually
getting at an idea which is very familiar and obvious. You can think of the Exclusion premise as
saying that every event has only one complete explanation. There are many particular facts that might go into explaining some event, but they will always add up to one, and only one complete explanation of the event.

Before looking at complicated problem cases, let’s just focus on how Exclusion is supposed to apply to mind-body causation. The idea is simply that, if there is a complete physical story about why I pushed the button, then there can’t be a complete but separate mental story about the same button-pushing. Conversely, if there is a mental story that completely explains the button-pushing, there can be a complete separate physical story about the same. Note that all parties to the current debate agree with this idea. Physicalists think there is a physical explanation for our actions, but no separate mental story. And interactionist dualists think there is a non-physical explanation for our actions, but no complete physical story. Some form of Exclusion is endorsed by nearly everyone.

Theories of mind and body aside, why believe Exclusion? I think the justification for Exclusion comes both from our day-to-day explanatory practice, but also from scientific practice. Suppose you come home and find your model batmobile broken. You think your roommate knocked it off the shelf. Your roommate says there was an earthquake which knocked it down. Both explanations are equally plausible. Now suppose that you find out there really was an earthquake, which really was powerful enough to knock down your batmobile. At this point, you will probably give up on the assumption that your roommate did it. Why? Well, because you already have a full explanation. And one full explanation is normally all there is. In other words, you are reasoning according to Exclusion.

Now for the complications: the official statement of Exclusion allows for some exceptions. It says that any given event has only one full cause, except in rare coincidental circumstances. These are precisely the cases of overdetermination discussed at the beginning, like the case of the two assassins who kill the same target at the same time. Similarly, it’s possible your roommate did knock the batmobile off the shelf, and at the exact same moment, a massive earthquake hit with the same effect. Still, this would be a remarkable coincidence, something that should occur only rarely.

In general, we can accept isolated instances of overdetermination. What is intolerable from the perspective of scientific explanation is systematic cases of overdetermination. If someone tells you that there is an earthquake every time your roommate knocks your toys off the shelf, you should be suspicious. Both stories can’t be right. Either the earthquaking isn’t happening, or the knocking isn’t happening, or (like that movie Colossal) the two are somehow part of the same process. Once again, Exclusion at work.

Let’s bring this back to the argument against dualism. According to Exclusion, no event has two full causes, but also that there are some exceptions, the rare cases of causal over-determination. Might the case of \( M \) (the mental cause of a physical event \( P \)) and \( P^* \) (the physical cause of \( P \)) be a rare case of causal over-determination?

I think the answer is pretty clearly no. For one thing, we constantly make decisions and take
actions, so downward causation is ubiquitous. At the same time, causal closure is supposed to apply universally. This would make the simultaneous occurrence of mental events and physical events utterly systematic, not isolated one-off incidents. This is exactly the kind of general pattern Exclusion is meant to address.

A second point here is that counting downward causation as overdetermination goes deeply against the interactionist dualist view of mental causation, and even the common sense view of mental causation. For the interactionist dualist, the whole point of downward causation is that the soul makes interventions into the physical world which would not have happened if physics had been left to its own course. If the force of every action of the soul were precisely reduplicated by a physical event in the body, this would no longer be the interactionist picture of causation. So the dualist won’t be pushing for this line of response to the Exclusion Argument.

As for common sense, we all assume that our actions depend on our decisions. If I had decided otherwise, my actions would have been otherwise. But this would be false if every case of downward causation were a case of overdetermination. For then, even if I had decided otherwise, the same action would have resulted from physics anyway. (Think of the assassins.) Carruthers puts the point this way: “a decision is sometimes causally necessary for a bodily movement to occur.” (Carruthers 2009, 154-155, my emphasis) In cases of overdetermination, neither full cause is an essential in provoking the result, since the other cause could always have handled it. But intuitively, our decisions are essential to provoking our actions.

Crane puts the same point this way:

It is nonetheless very implausible to suppose that this is the way mental and physical causes relate to each other. For if they did, then every time a mental state had some effect in the physical world, the completeness of physics guarantees that there would be a cause in one’s brain which is itself enough to bring about that very same effect. It then looks like a coincidence that my body manages to co-ordinate so well, given these distinct causes of its motions. Yet the idea that one’s control of one’s body is in this way coincidental is in conflict with everything we know about the causation of behaviour—either from our own experience, or from common sense. (Crane 2001, 50)

Long story short: I think that viewing downward causation as a case of overdetermination is kind of a non-starter. Physicalists don’t like that idea. Dualists don’t like that idea. I think it is better to think of Exclusion, working in the background, as simply ruling out simultaneous but separate mental and physical explanations of the same physical events.

6 Evaluating the Premises

How controversial is the Exclusion Argument? Which of its premises require further defense?

I consider Downward Causation and Causal Closure to be the most important and interesting
premises of the argument. Each reflects a significant view about how the world works, and each
has been widely debated by philosophers. Interactionist Dualists, of course, reject Causal Closure,
so there is a substantive empirical question about whether physical events all have physical causes.
And later in the quarter, we’ll see that other kinds of dualists (not interactionists) reject Downward
Causation. So there is again a substantive metaphysical question about whether or minds really do
affect our bodies in the way that Downward Causation suggests.

On the other hand, Exclusion seems to me to be a minor or supporting premise. It is essential to the
argument, but it basically serves to make a kind of uncontroversial background point completely
explicit. It can seem confusing, because it is so abstract, but most people agree that it is by-and-large
correct. It is more like an axiom or metaphysical truism, and less like a substantive claim. Debate
about whether downward causation might be a case of overdetermination is mostly a distraction
from the key issue, since both interactionist dualists and physicalists firmly count mental causation
as a standard, non-overdetermined case of causation.

References

Papineau, David (2009). “The causal closure of the physical and naturalism”. In: *The Oxford hand-
book of philosophy of mind*. 

REFERENCES 9