Problems with Causal Exclusion

Ana Diaz
Montana State University Billings

Physicalism has played an essential role in our understanding of the mind and its place in the world throughout the twentieth century. It is considered by many to be beyond reproach and the predominant view in philosophical thinking. The assumption generally made is that the world is physical and does not contain, as Descartes believed, immaterial, non-extended substances. Dualism has therefore been roundly discarded and supplanted by physicalism as the foundational ontological position that has within its discourse talk of fundamental physical properties that form the bedrock of reality. Not surprisingly, this position has configured much of the debates and problems in philosophy of mind where understanding the place of mentality in a physical world is a central project. Since mental properties are not fundamentally physical ones, the challenge has been to find a way of accommodating the mind while at the same time upholding the truth of physicalism.

The difficulties of maintaining a physicalist outlook while at the same time admitting non-physical properties have become too great to overlook. Among the problems facing the non-reductive physicalist is articulating a view of how the mind can exert any causal powers in a causally closed physical world. If causation can only occur within the physical domain, or as is often believed, between fundamental physical items, then mental properties look to be causally inert and not responsible for the behaviors often attributed to them. Physicalism appears to render our mental lives epiphenomenal at best. At worst, if our mental lives make no causal difference, then one has to wonder whether there really are any mental properties at all. Ultimately what is at stake in this debate is a version of Mental Realism whereby our mental states—beliefs, desires and intentions—in voluntary action appear in the causal chains that lead to the movement of our limbs and the rearrangement of our physical world. As Fodor puts it, if we cannot make sense of the mental as *literally* causal, "that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species; if we're wrong about the mind, then that's the wrongest we've ever been about anything."[1]

Although the non-reductive view has maintained a significant amount of popu-

larity, albeit amongst a string of criticisms since its surge in the early 1970's, the pendulum has begun to swing back to reductionism. The belief is that the demands of physicalism require that the mind be reduced to fundamental features of the physical world. No one has done more to spearhead this movement than Jaegwon Kim, a zealous voice for reductive physicalism. Once a defender of the supervenience relation and the non-reductive view, he now claims that these positions make little sense and with them, the notion that the mind is causal at all. He argues, with the publication of *Mind in a Physical World* and *Physicalism, or Something Near Enough*, that the non-reductive position cannot be made consistent with the constraints of physicalism.[2] Therefore, committing to a physicalist outlook gives rise to the problem of mental causation, which suggests that the mind cannot be construed as having the causal efficacy to exert any powers in a fundamentally physical world, despite our pretheoretical understanding of the mind.

According to Kim (and others,[3]) there are a set of principles he believes any serious physicalists must hold, even non-reductive ones, that conspire to make problems for the mental. They are the causal closure of the physical domain, the principle of causal exclusion, and the principle of causal overdetermination. Acceptance of these results in making the mind causally impotent and as such puts the onus on the non-reductive physicalist to explain how mental causation is possible in light of these constraints.

In this paper I want to focus on the principle of causal exclusion, which conceptually entails the precept that effects are not systematically overdetermined. While this is an intuitive assumption to make about causes and their effects, causal exclusion rests on the idea that there are sufficient causes that strip all other causes from doing any work in the production of an effect. The main worry is in avoiding overdetermined consequences, and causal exclusion as it is generally understood seems to do just that. However, it does so at a cost. In systematically avoiding causal overdetermination, causal exclusion also makes some causes of an effect obsolete, creating a puzzling paradigm. Rather than solving (or explaining away) the problem of overdetermination, causal exclusion in fact creates a paradox about causation. In some contexts, as I go on to show, causal exclusion results in certain seemingly necessary causes being inert. Without invoking these causes, it becomes perplexing how the result is reached. I begin by articulating these two causal principles and the manner in which both make the mental causally impotent.

There are two metaphysical precepts that are believed to be necessary for understanding causation. Both are meant to express conceptual truths that should be universally accepted by anyone considering these matters. The first is the principle of causal overdetermination as construed by Kim, which states that:

> It would be unreasonable for us to suppose that all physical events are the result of genuine cases of causal overdetermination, where independent events *f* and *g* are both sufficient for the production of some other event *e*.

According to this principle, it is highly implausible that all physical events have two independent causal chains that lead to the very same effect. If the world were overdetermined in just this way, then its causal structure would be like the example of the

assassin who fatally shoots his victim while the victim at the same time suffers a fatal heart attack. It is highly improbable that such cases permeate and are prevalent in the causal structure of the physical.

The other related principle on causal exclusion as expressed by Kim states that "If an event *e* has a sufficient cause *c* at *t*, no event at *t* distinct from *c* can be a cause of *e* (unless this is a genuine case of causal overdetermination)."[4] The intuitive idea is that if there are two purported sufficient causes of a single event, two causal stories both of which are meant to offer a full account of that event, then only one of those causes will be sufficient for the event's production, unless it is a genuine case of causal overdetermination. The notion of causal sufficiency therefore states that if *c* is causally sufficient for *e*, then if *c* is instantiated, *e* must be also, irrespective of whether any other property is instantiated (or not) at that time. Since causal overdetermination is not ubiquitous, as suggested above, for effects where there are two alleged causes, only one counts as sufficient to produce the effect. The other cause, in such instances, is robbed, as it were, of all its causal powers since the effect would have been produced regardless of its appearance. It is, in short, epiphenomenal.[5]

The picture that emerges is akin to Block's worry about what exactly causes the bull's anger, and are linked to the concern that there might be an over abundance of causes.[6] Is the bull's anger the result of the red cape or its provocativeness? Counting the cape's color and its provocativeness as a cause of the bull's anger seemingly leads to causal overdetermination since the bull's anger is the effect of two distinct causes. Kim does not see this as a serious worry. He says that "if the color of the cape is, in and of itself, a sufficient cause of the anger,...what *further* causal work is left for its provocativeness?"[7] Since by stipulation this is not a causally overdetermined case, the principle of causal exclusion indicates that the property of provocativeness makes no special contributions to the bull's anger. While it may appear as though the cape's provocativeness acts as a cause, this turns out to be nothing more than a pseudo cause on a Kimean analysis.

Worries about mental causation are akin to Block's concerns over what property is involved in causing the bull's anger. Given a layered conception of properties, it may appear that mental property M is the cause of mental property M*. However, M* is realized by P*, which leaves us with two supposed complete and independent causes of M*, namely, M and P*. Except for isolated cases, we cannot assume M* to be overdetermined since this construal would render as fictional all voluntary behavior. If this is not a genuine case of causal overdetermination, then M* has one sufficient cause. Regardless of the causal work we may think or want M to play, P* appears to be a completely sufficient cause of M* since M* would not make an appearance without P* being instantiated. This means that M is causally preempted from playing any role in the production of M*.

The upshot of all this is that no matter how we slice it, mental properties look to be causally impotent in the production of any event. The picture that emerges—to use one of Kim's examples—is similar to "a series of shadows cast by a moving car: there is no causal connection between the shadow of the car at one instance and its shadow an instant later, each being an effect of the moving car."[8] In this analogy, the car represents genuine causal processes while the shadows exemplify what we might call

pseudo causal processes, which by all appearances are law-like and regular. Similarly, while the mental may seem to be law-like and regular, it is merely an appearance. *Real* causal work in fact occurs in an underlying stratum.

I want to consider whether the notion of a sufficient cause, which underpins the principle of causal exclusion, makes sense. The idea is that if *c* is a sufficient cause of *e*, then the presence of *c* necessarily results in the presence of *e*. Otherwise it is a case of causal overdetermination. Let us consider an example outside the domain of the mind/body problem to test the soundness of causal sufficiency. It is commonly assumed that there is a genuine difference between the hardware and software of a computer. The hardware is that physical stuff that I can come into contact with and which occupies a certain amount of space. Software, by distinction, is considered to be the non-tangible component of a computer that can be instantiated in some piece of hardware. However, the software itself is not identical to what realizes it.

While this distinction is commonly accepted, it is not universally considered to be the case that software is distinct from hardware.[9] I want to briefly make an argument as to why software is not identical to hardware, even if it is true that software at a particular time *t* gets physically implemented vis-à-vis a series of commands into the hardware. First, a software program is a set of instructions actualized in the physical structures of a computing device. The encoding that makes up the software program represents a computational function. So, for example, Word is a software program that is encoded differently depending on the physical device upon which it is found. So, Word for Mac is the same software program as Word for Windows but encoded differently—that is, each set of instructions is distinct. Second, software is a functional kind of entity, even though instances of code realize it. The function of Word is for it to behave as a word processor. While some instances of code realize Word, the software, *qua* a function, is not identical to the particular encoding since program code varies. In other words, while my Mac has Word encoded into it through a series of binary numbers, that is not identical to the function that code represents given the multiple ways Word can be realized.[10] Moreover, even when my computer is turned off, Word or the functions it represents, do not cease to exist.

Given this distinction, let us imagine that I am currently using Word to write this paper and every time I press the "return" (or "enter") key, a carriage return occurs. This appears on the screen as a hard line break with the cursor moving to the left margin. I may reasonably wonder as to the cause of the carriage return. The obvious candidates are the physical structures of the computer along with the program code as it is implemented in my device and the software itself, which dictates how the keyboard functions. On the one hand, it appears that the reason the enter key functions as it does with Word is entirely because of the software program. If I were to open Chrome or Firefox, type a URL and press the return key, an entirely different function would be performed. The function of creating a line break when the enter key is pressed is what causes the cursor to move in a particular way. On this account, the software appears to be a sufficient cause for why the return key performs the function of a carriage return when pressed.

On the other hand, the reason for a hard line break when I press enter is entirely accounted for by the physical structures of my device. There is the physical encod-

ing that instructs the machine to perform a carriage return when I press enter. And, as Turing reminds us, computers are input/output devices that cannot alter their states without some physical input that acts upon the physical states of the machine. It is my physical act of pressing enter that physically alters the state of my Mac. This, too, looks like a sufficient cause for the hard line break.

There are apparently two causal stories here, both of which look like sufficient causes as to why the carriage break occurs. If this is right, then we have entered the domain of causal overdetermination, in which case my computer turns out to be systematically overdetermined. But, as the principles above suggest, this cannot be the case. Causal exclusion tells us that only one cause is sufficient for the effect while the other is *entirely* inefficacious. I want to consider how paradoxical the results are when causal exclusion is applied in this context.

On the one hand, we have to say that the reason for the hard line break when I press the enter key is the result of the software alone. This is not the same as claiming that the software works in tandem with the hardware to produce the effect. Rather, it is the claim that the software *alone* accounts for the effect while the hardware plays no causal role. Conversely, one could say that the sufficient cause of the carriage return is the hardware, the actual physical encoding and structure of the device. While this may seem like an attractive option, the physical encoding is not the same as the software. The encoding merely represents a function. As William Duncan suggests:

> The computational function is an entity, the actualization of which (speaking loosely) "does" something. The "doing" is, of course, an occurrent. It unfolds in time. However, the function, itself, is not a temporal entity. It exists even when the function is not being actualized.[11]

If this is right, then causal exclusion results in the view that the function of the software plays no role in the production of a line break.

There is something puzzling about the picture that emerges above, something that is no doubt closely analogized with the mind/body problem. However, as far as I know, there is no analogous hardware/software problem. Certainly, there does not appear to be a concern over the causal efficacy of software. Erase them from your laptop and the device simply becomes an expensive tray. And yet, applying causal exclusion to this example creates a problem of causation where none seems to have existed before. In a sense, causal exclusion creates a false choice between either having the hardware alone do all the causal work or the software. By suggesting that only one of these counts as the sufficient cause, the other need not have even made an appearance. But perhaps this is not the most fruitful way of thinking about causation in every context. I am not claiming that causal exclusion is wrong, *per se*, but simply that it may not be as obviously true as Kim, amongst others, supposes it to be.

Notes

1. *Sic.* J. Fodor (1998) pg. xii.
2. Kim (1998, 2005)
3. Malcolm (1968), Antony and Levine (1999), Papineau (2000), McLaughlin (2006),

4. Kim (2005) pg. 17. This principle is also argued for in Kim (1998) .
5. I should note that this principle does not favor the physical over the mental. It is silent with regard to which counts as a sufficient cause. As such, causal exclusion alone does not allow one to conclude that the mental is causally inefficacious. However, causal exclusion together with causal closure of the physical, does make mental properties causally inert.
6. Block (1990)
7. Kim (1998) pg. 53.
8. Kim (1998) pg. 45.
9. See Moor (1978) and Suber (2009)
10. For more on this topic, see Duncan (2009). Here he also has a fruitful example from a non-computational context. If the function of a claw hammer is to drive nails and pull them, this function does not cease to exist when the hammer is not in use. Rather, Duncan claims, the hammer always possesses this function, even when it's not actualized. In this sense, the function is not a temporal physical process since it exists even when not instantiated.
11. Duncan (2009) pg. 22.

Works Cited

Antony, Louise and Joseph. Levine, 1997, "Reduction With Autonomy," *Philosophical Perspectives*, 11, pg. 83–105.

Block, Ned (1990). "Can the Mind Change the World?" in *Meaning and Method: Essays in Honor of Hilary Putnam*, pg. 137-170.

Duncan, William (2009). "Making Ontological Sense of Hardware and Software," <http://www.cse.buffalo.edu/~rapaport/584/S10/duncan09-HWSWOnt.pdf> (accessed January 14, 2014).

Kim, Jaegwon (1998). *Mind in a Physical World*. Cambridge, MA: MIT Press.

Kim, Jaegwon (2005). *Physicalism or Something Near Enough*.

Malcolm, Norman (1968): "The Conceivability of Mechanism," *Philosophical Review*, 77, pg. 45–72.

McLaughlin, Brian (2006). "Is Role-Functionalism Committed to Epiphenomenalism?," *Journal of Consciousness Studies*, 13, pg. 39–66.

Moor, James (1978). "Three Myths of Computer Science," *British Journal for the Philosophy of Science* 29:3, pg. 213-222.

Papineau, David (2000). "The Rise of Physicalism," in *Physicalism and Its Discontents*, pg. 3-36.

Suber, Peter (1988). "What is Software?," *Journal of Speculative Philosophy* 2:2, pg. 89-119.