BRILL

Causal Exclusion and Causal Autonomism

Author(s): Weixin CAI

Source: *Frontiers of Philosophy in China*, September 2018, Vol. 13, No. 3 (September 2018), pp. 402–419

Published by: Brill

Stable URL: https://www.jstor.org/stable/10.2307/27073802

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/10.2307/27073802?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

CAI Weixin

Causal Exclusion and Causal Autonomism

Abstract   The causal exclusion problem is often considered as one of the major difficulties for which non-reductive physicalists have no easy solution to offer. Some non-reductive physicalists address this problem by arguing that mental properties are to some extent causally autonomous. If this is the case, then mental properties will not be causally excluded by their physical realizers because causation, in general, is a relation between properties of the same level. In this paper, I argue that the response from causal autonomy cannot be successful for two reasons. First, it does not offer a satisfactory explanation for how mental particulars can have causal efficacy in a non-reductive physicalist framework. Second, the causal considerations underpinning this response do not really support the conclusion that mental properties are causally autonomous.

Keywords   non-reductive physicalism, causal exclusion, causal autonomism, interventionism, mental particulars, token causatio, propostionality requirement

Physicalism is usually formulated in terms of *the supervenience thesis*—the claim that mental properties supervene on physical properties. That is, a system can instantiate a mental property only if it instantiates some physical property that underlies this mental property. Given that supervenience captures only the *logical* relationship between the occurrence of mental properties and the occurrence of physical properties, physicalists are intrigued by the question of what *metaphysical* relationship between mental properties and physical properties grounds their supervenience relationship. More exactly, physicalists disagree about whether a mental property can be identical with, or reduced to, its supervenience base. Understanding the mental-physical supervenience relation as grounded by a kind of metaphysical dependence relation, non-reductive physicalists hold a negative answer to this question: a mental property can be realized by different physical properties and thus cannot be identified with or reduced to any of its physical realizers. Jaegwon Kim (1998; 2005) challenges

CAI Weixin (✉)
Department of Philosophy, Simon Fraser University, Burnaby, V5A 1S6, Canada
E-mail: weixinc@sfu.ca

non-reductive physicalism with the causal exclusion argument: If mental properties are not identical with the physical properties on which they supervene, they will be causally impotent because their causal work will be taken over by their physical realizers. Thus, Kim concludes that a physicalist must endorse a reductive form of physicalism to maintain the causal efficacy of mental properties.

Some non-reductive physicalists address the causal exclusion problem by arguing that higher-level properties, such as mental properties, are more or less causally autonomous, even though they metaphysically depend on physical properties. The causal autonomy of mental properties guarantees their causal efficacy in a non-reductive physicalist framework, because they do not actually compete with their physical realizers (Yablo 1992; List and Menzies 2009; Campbell 2010a; Campbell 2010b; Zhong 2014). I will refer to this position as "causal autonomism" in the rest of the paper. As a typical argumentative move, causal autonomists try to show that it is generally the case that causal relationships are formed between properties of the same level. That is, normally, a mental cause has a mental effect and a physical cause has a physical effect. Therefore, when a mental property causes a mental effect, there is no ground to infer that its physical realizer also causes the same mental effect. No causal exclusion can emerge in this picture.

In defense of their view, one major line of argument chosen by causal autonomists is to show that the fact of causal autonomy naturally follows from a widely accepted conception of causation which is often expressed as the claim that c causes e if and only if a change in c will lead to a difference to e. They believe that this so-called "difference-making" conception of causation sets a constraint on causal relations, which require that causes must be proportional to their effects.[1] I will argue that this line of argument cannot successfully address the causal exclusion argument. In the first two sections, I briefly recap the causal exclusion argument and show how causal autonomists address this by invoking the proportionality constraint on causation. Then I offer two independent arguments against their responses: One aims to show that their responses cannot secure the causal efficacy of mental particulars, and the other aims to show that

---

[1] Yablo (1992), List and Menzies (2009), and Zhong (2014) favor this line of argument. In contrast, Campbell (2010a; 2010b) argues for the causal autonomy of mental properties by appealing to the requirement that no two variables whose relationship is logically or metaphysically necessary should be contained in the same causal system. Since there are intimate metaphysical connections between mental properties and their physical realizers, this requirement, if it works, can exclude physical variables as appropriate candidate causes in any causal system where the mental variables they realize exist. I will not discuss this line of causal autonomism in this paper; however, if my argument is correct, it may serve as evidence that Campbell's strategy is more favorable to causal autonomists, compared with the line of argument I am to argue against here.

there is no plausible reason to favor their formulations of the difference-making account of causation over other alternatives. For the sake of simplicity, I will primarily focus on Lei Zhong's argument for causal autonomism in my discussion. But my critiques can also be applied to other causal autonomists who argue in a similar way.

1   The Causal Exclusion Argument

With the causal exclusion argument, Kim (1998; 2005) tries to show that non-reductive physicalism is not an appealing position for physicalists because irreducible mental properties cannot maintain their putative causal powers in a physicalist framework. Broadly speaking, this argument relies on the following premises.

Supervenience: Mental properties supervene on physical properties.
Irreducibility: Mental properties are not reducible to, and are not identical with, physical properties.
Closure: Every physical event has a (sufficient) physical cause.
Exclusion: No single event can have more than one sufficient cause at the same time, unless it is a genuine case of overdetermination.
Non-overdetermination: In the case of mental causation, a mental property and its subvenient physical property do not overdetermine the effect.

The idea is that if we accept non-reductive physicalism (i.e., the conjunction of *Supervenience* and *Irreducibility*), a higher-level mental property and its lower-level physical realizer will be two distinct causes of the same effect (for simplicity, I use "a cause" to refer to a sufficient cause). To see this, suppose a mental property, M_1, supervenes on but is not identical with a physical property, P_1. Further suppose M_1 causes another physical property, P_2. According to *Closure*, P_2 has a physical cause which is presumably P_1. So M_1 and P_1 are supposed to be two distinct causes of P_2. However, given *Exclusion* and *Nonoverdetermination*, P_2 can only have a single cause. Since M_1 and P_1 are not identical, one of them must be excluded from being a cause of P_2. To stick with *Closure*, as all physicalists should, we are pushed to choose P_1 over M_1 to be the cause of P_2. Otherwise, it will lead to an infinite regress until some physical property is chosen as the cause of P_2. Thus, Kim concludes that if non-reductive physicalism is true, the putative causal powers of mental properties will be systematically excluded by those of their physical realizers *as long as* mental properties are supposed to cause the same effects as their physical realizers.

Non-reductive physicalists can grant that if this argument works, it demonstrates at best that irreducible mental properties have no power to cause

physical effects. But this by no means entails that irreducible mental properties are causally impotent *simpliciter*, since mental properties can still be causally efficacious in bringing about mental properties or other kinds of higher-level properties.[2] Thus, even though M_1 does not cause P_2, it may still cause M_2. In order to argue that M_1 is causally impotent in this regard, reductive physicalists must show that when M_1 causes M_2, P_1 also causes M_2. They also need to show that when both M_1 and P_1 are two purported causes of M_2, M_1 will be excluded by P_1 from being a cause of M_2. However, none of the principles invoked by the type of causal exclusion argument we have seen can establish these claims. Thus, if non-reductive physicalists are only concerned with preserving mental-mental causation or the causal relations between mental properties and some other kinds of higher-level properties, their view will not be threatened by this type of causal exclusion argument.

Another type of causal exclusion problem (what Zhong calls "sophisticated exclusion") aims to show that even if non-reductive physicalists restrict the causal efficacy of mental properties to higher-level properties, causal exclusion still occurs. This type of causal exclusion argument may appeal to a causal principle that posits a causal connection between P_1 and M_2, such as the following one indicated by Zhong:

> **Causal Realization**: If property X causes property Y, then either **Upward Causation** X causes any higher-level property of Y to be instantiated on this occasion, or **Downward Causation** X causes any underlying property of Y to be instantiated on this occasion. (Zhong 2014, 347)

---

[2] It is noteworthy that the set of higher-order properties which mental properties are purported to cause include actional properties. Since an action can be realized by a variety of bodily movements, actional properties are higher-order properties supervening on lower-level physical properties of which concrete bodily movements consist. This distinction is usually helpful to reveal the causal structure in cases like the one below. In an example discussed by both Woodward (2008) and List and Menzies (2009), the neuroscientist Richard Andersen and his colleagues discovered that there was a strong correlation between monkeys' intentions to reach certain goals and the corresponding actions they performed. In this experimental setting, each of the monkeys' intentions is identified with a unique aggregate pattern of neuron firings, while each pattern can be realized by different configurations of individual neurons. The example mentions three kinds of properties: an intention, I_1, a specific configuration of the behavior of individual neurons, N_1, which realizes the aggregate pattern of neuron firings associated with I_1 and thus can also be viewed as a lower-level realizer of I_1, and an action, A_1, which follows from I_1. Given the strong correlation between I_1 and A_1, it is tempting to attribute a causal relation between them, instead of between N_1 and A_1. But now the causal structure here is incomplete, for it does not specify what the effect of N_1 is. A failure to address this might threaten *Closure*. Once we remind ourselves of the fact that A_1, as an action, can be realized by different bodily movements, we can fill in the causal picture with a lower-level physical property, B_1, which realizes A_1. We are then able to claim that *Closure* is not violated because it is B_1, rather than A_1, that is the lower-level physical effect of N_1 in the current causal structure.

For the sake of argument, suppose M_1 causes another mental property, M_2, on some occasion. Also, suppose M_2 is realized by P_2 on this occasion. According to *Downward Causation*, if M_1 causes M_2, M_1 must also cause P_2. But given *Closure*, P_1 causes P_2 as well. So M_1 and P_1 are two distinct causes of P_2. This leads to the same kind of causal exclusion we saw earlier. Consequently, M_1 cannot be a cause of P_2, which, by *modus tollens*, entails that M_1 cannot be a cause of M_2.[3] Based on this line of reasoning, a reductive physicalist can conclude that it is impossible for an irreducible mental property to cause either a mental effect or a physical effect. That is, non-reductive physicalism leads to epiphenomenalism. If this is the case, we have strong reason to reject non-reductive physicalism.

2   The Argument for Causal Autonomism

Given that causal considerations play a central role in the elaboration of the causal exclusion argument, it is not accidental that many non-reductive physicalists attempt to reject this argument by appealing to some independently reliable accounts of causation. More specifically, many of them invoke interventionism to examine whether certain properties in the standard causal structure introduced by the causal exclusion argument are causally connected (e.g. Shapiro and Sober 2007; Woodward 2008; Woodward 2015; Campbell 2010a; Campbell 2010b; Zhong 2014). Interventionism is usually regarded as one of the most plausible accounts of causation nowadays and has been widely endorsed in scientific investigations.[4] The motivation behind interventionism is the idea that a cause makes a difference to its effect. Interventionists believe that this is an intrinsic feature of any causal relation and try to explicate the nature of causation in terms of how the effect will respond when the cause is properly manipulated in some way.

For example, Woodward offers an interventionist account of causation as follows.

M: A necessary and sufficient condition for X to be a (type-level) direct cause of Y with respect to a variable set **V** is that there be a possible intervention on X that will change Y or the probability distribution of Y when one holds fixed at some value all other variables Z_i in **V**. A necessary and sufficient condition

---

[3] There is a parallel argument to the same conclusion by appealing to *Upward Causation*. Because it is structurally similar with the argument from *Downward Causation*, I will not discuss *Upward Causation* in the remainder of this paper. It is easy for the reader to adapt my discussion to the argument from *Upward Causation* with some modest modifications.
[4] See Woodward (2003) for the most philosophically developed formalization and defense of interventionism.

for X to be a (type-level) contributing cause of Y with respect to variable set $\mathbf{V}$ is that (i) there be a directed path from X to Y such that each link in this path is a direct causal relationship, and that (ii) there be some intervention on X that will change Y when all other variables in $\mathbf{V}$ that are not on this path are fixed at some value. (Woodward 2003, 59)

Woodward's distinction between direct cause and contributing cause comes from his view that there are different kinds of causal relationships that come under the same name, "cause." Our ordinary conception of causation can thus be characterized as: X causes Y if and only if X is either a direct cause or a contributing cause of Y. As a received practice for interventionists, Woodward constructs the causal relata as variables which can have different values and hence allow for being intervened on from one value to another. For instance, the variable "age" can have "0–18 years old," "19–45 years old," "46–70 years old," and "more than 70 years old" as its different values. A possible intervention on age will change its value from one to another. It is worth noting that according to this interventionist account of causation, in order to establish the causal connection between age and cognitive abilities, we do not need to prove that *every* possible intervention on age should be accompanied with a change in cognitive abilities. This criterion is often viewed by many interventionists as not only unnecessarily strong but hard, if not impossible, to follow in practice. Thus, given Woodward's interventionism, there is a causal relationship between age and cognitive abilities if and only if there is *some* possible intervention on age alone that can make a change in cognitive abilities, when other relevant factors are held fixed.

Property causation is a specific kind of causal relation where the relata are variables having "being present" and "being absent" as their only two values. We can have an account of property causation in the same spirit of Woodward's interventionist account.

> **Property Causation₁:** Property X causes property Y if and only if there is a possible intervention on X such that if that intervention were to occur, the value of Y or the probability distribution of Y would change when other relevant variables are held fixed.

Using "p" and "a" to stand for a property's being present and its being absent respectively, X causes Y if and only if when some possible intervention changes X's value from $X = x_p$ to $X = x_a$ (or vice versa), a corresponding change will occur to Y.[5]

---

[5] Woodward (2015) employs his interventionism to address the causal exclusion argument. Unlike causal autonomists who want to show that there is no mental-physical causation in general, Woodward argues that it is *Non-overdetermination* that should be rejected because, according to interventionism, both an irreducible mental property and its physical realizer can be causally related with the same effect.

However, most causal autonomists see Woodward's interventionism as too weak to characterize the difference-making conception of causation accurately. From their perspective, it is insufficient to show that two properties are causally connected if what we know about them is merely that by manipulating one property we can *sometimes* manipulate the other. Rather, a manipulation on the cause should *always* be followed by a change in its effect. Maybe this is because causation is often conceived as a kind of necessity: When a cause occurs, the causal laws instantiated by this cause guarantee a certain effect to occur. If it is only the case that by manipulating a cause from being absent to being present (or vice versa) a corresponding change occurs on its effect in some circumstances but not in others, then it is unclear how the cause can be said to necessitate its effect. Therefore, the idea of causal necessitation seems to suggest that a cause must universally co-vary with its effect in some sense.

Moreover, causal autonomists claim that our causal intuitions are backed up by the idea that a cause makes a difference to its effect by universally co-varying with it. They usually appeal to Yablo's pigeon example (1992) as an illustration. Suppose a pigeon is trained to attack red and only red objects presented to it. Consequently, when a scarlet object is presented, the pigeon will attack this object. When reflecting upon whether it is the redness or the scarletness of the object that causes the pigeon to attack, many of us are inclined to assert that the redness of the object is the genuine cause. Given that the major difference between redness and scarletness is their specificity, the most reliable explanation of our intuitions seems to be this: Because the pigeon might still attack when some non-scarlet object is presented, the scarletness of this object is too specific to pick out what is really causally responsible for the pigeon's attack. In contrast, it is more appropriate to regard redness as the cause, for this implies that when a non-red object is presented, the pigeon will not attack, which nicely fits the facts about the pigeon. Thus, causal autonomists claim that it is embedded in the folk conception of causation that a cause should be *proportional* to its effect. That is, the presence of a cause should be stably accompanied with the presence of its effect, and the absence of a cause should be stably accompanied with the absence of its effect. For this reason, causal autonomists tend to advocate an account of causation which explicitly characterizes the proportionality requirement of the cause to its effect.[6] For example, Zhong endorses an account of property causation as follows.

Property Causation₂: Property X causes property Y if and only if:
(N1) If an intervention that sets $X = x_p$ were to occur (while all other relevant variables in the causal graphs are fixed), then $Y = y_p$; and

---

[6] In contrast, Woodward (2010) sees proportionality as a virtue for some causal relationships over others, rather than as a requirement on every causal relationship.

(N2) If an intervention that sets $X = x_a$ were to occur (while all other relevant variables in the causal graphs are fixed), then $Y = y_a$. (Zhong 2014, 344)

Note that it is not essential for an account of causation that accords with the proportionality requirement to be formulated in interventionist terms. For instance, List and Menzies formulates an alternative account of property causation in terms of possible-world semantics:

Property Causation₃: *F* causes *G* if and only if the presence of *F* makes a difference to the presence of *G* in the actual world if and only if it is true in the actual world that
(i) *F* is present $\square\rightarrow$ *G* is present; and
(ii) *F* is absent $\square\rightarrow$ *G* is absent. (List and Menzies 2009, 483)

Despite the difference in modal commitments, these causal autonomists stick with the very same idea in their accounts of property causation: A cause must co-vary with its effect in a highly stable manner. They then employ these accounts to examine whether mental properties and physical properties are causally connected in the causal structure that emerges out of the causal exclusion argument. Since these causal autonomists adopt similar arguments for their claims, I will illustrate how they reject the causal exclusion argument with a primary focus on Zhong's argument.

Compared with the existence of inter-level causal relationships such as causal relationships between mental properties and physical properties, it is much less controversial that there are intra-level causal relationships if anything is causal. It is thus commonly accepted that in the causal structure introduced by the causal exclusion argument, we can safely begin with the assumptions that M_1 causes M_2 and that P_1 causes P_2. It is a principle like *Causal Realization* that plays a central role in inferring inter-level causal relationships (e.g. M_1 causes P_2) from intra-level causal relationships (e.g. M_1 causes M_2). This move is crucial because it is employed by reductive physicalists to show that when irreducible mental properties do not cause physical properties (as a result of causal exclusion), they also do not cause mental properties. Thus, if non-reductive physicalists can reject *Causal Realization*, they are able to break the intimate ties between intra-level and inter-level causal relationships: When M_1 causes M_2, it need not also cause P_2. Thus, the potential causal competition between mental properties and their physical realizers can never arise. The resultant picture is that there are only intra-level causal relationships in most cases. In this way, the causal efficacy of mental properties can still be maintained by non-reductive physicalists. This is the strategy Zhong adopts in addressing the causal exclusion argument.

Causal Realization (*Downward Causation*) entails that when M_1 causes M_2, M_1 also causes M_2's supervenience base on that occasion, i.e. P_2. According to

Property Causation$_2$, M_1 causes M_2, if and only if:

> (1) If an intervention that makes M_1 present were to occur (while all other relevant variables are fixed), then M_2 would also be present;
> (2) If an intervention that makes M_1 absent were to occur (while all other relevant variables are fixed), then M_2 would also be absent.
> Similarly, M_1 causes P_2, if and only if:
> (3) If an intervention that makes M_1 present were to occur (while all other relevant variables are fixed), then P_2 would also be present;
> (4) If an intervention that makes M_1 absent were to occur (while all other relevant variables are fixed), then P_2 would also be absent.

Causal Realization thus entails that when (1) and (2) are true, (3) and (4) must also be true. It is then possible to verify whether *Causal Realization* is a reliable principle by considering whether the truth of (1) and (2) guarantees the truth of (3) and (4).

Zhong, along with most non-reductive physicalists, accepts the multiple realizability of mental properties: a single mental property can be realized by different physical properties.[7] Given the multiple realizability of mental properties, suppose M_1 is realized by P_1 on some occasions and realized by P_1* on other occasions. Also suppose that when M_1 is realized by P_1, its effect, M_2, is realized by P_2 and that when M_1 is realized by P_1*, M_2 is realized by P_2*. Zhong argues that while (1) and (2) are true, (3) might not be true because when there is an intervention making M_1 and hence M_2 present on some occasion, it is possible that M_2 is realized by P_2* rather than P_2. Thus, according to *Property Causation$_2$*, the fact that M_1 causes M_2 does not entail that M_1 causes P_2, which goes against what *Causal Realization* (*Downward Causation*) predicts.[8] This can serve as a decisive reason to reject *Causal Realization*, if we accept that *Property Causation$_2$* is a plausible account of property causation. Since the causal exclusion argument presumes a wrong-headed causal principle, Zhong concludes that it is unable to pose any threat to non-reductive physicalism. Once physicalists recognize that most causal relationships are intra-level ones, which exist between properties of the same level, they can attribute causal powers to higher-level mental properties without worrying that doing so will lead to causal conflicts with lower-level physical properties.

I have introduced Zhong's argument for the causal efficacy of mental properties from causal autonomism. If this approach is successful, then when a

---

[7] See Putnam (1967) and Fodor (1973) for classic formulations of multiple realizability.
[8] In a similar way, Zhong also argues that *Upward Causation* should be rejected because according to *Property Causation$_2$*, the fact that P_1 causes P_2 does not by itself entail that P_1 causes M_2.

mental property causes some mental effect, it does not cause the lower-level physical realizer of this mental effect; and when a physical property causes some physical effect, it does not cause the higher-order mental property realized by this physical effect. Causal relationships thus obtain between properties at the same level, rather than properties across different levels, in most cases. Compared with reductive physicalism, non-reductive physicalism is an equally appealing, if not better, option for physicalists. However, in the rest of the paper, I will present two independent reasons to hold a skeptical view on causal autonomism.

3   The Causal Impotence of Mental Particulars

A mental property is usually regarded as a type which can be instantiated on different occasions. There are type-causal relationships whose causal relata are properties as well as token-causal relationships whose causal relata are particulars such as events or property instances. Correspondingly, there are two ways of making sense of the question of whether the mental can be causally efficacious when it is not identical with the physical. On one hand, it might be asked whether M_1, as a mental property, can have any causal power over and above, and hence not excluded by, P_1, given that M_1 and P_1 are not the same *property*. The kind of causal exclusion argument employed by reductive physicalists aims to show that M_1 is causally impotent in this case. Call it *the type-causal exclusion argument*.

On the other hand, the question might be asked whether m_1, which is an instance of M_1's being realized by P_1 *on this occasion*, can have any causal power not excluded by p_1, which is an instance of P_1 that realizes M_1 on this occasion and subvenes m_1, given that m_1 and p_1 are not the same *property instance*. Although the causal exclusion argument normally concerns the causal efficacy of irreducible mental properties, it is easy to construct a parallel argument which throws doubt on the causal efficacy of irreducible mental particulars. For example, we can modify the premises of the type-causal exclusion argument as follows.

Supervenience*: On every occasion, mental property instances supervene on physical property instances.
Irreducibility*: On every occasion, mental property instances are not reducible to, and are not identical with, physical property instances.
Closure: Every physical event has a (sufficient) physical cause.
Exclusion: No single event can have more than one sufficient cause at the same time, unless it is a genuine case of overdetermination.
Non-overdetermination*: In the case of mental causation, a mental property instance and its subvenient physical property instance do not overdetermine

the effect.

Causal Realization*: If property instance x causes property instance y, then either **Upward Causation*** x causes any higher-level property instance of y to be instantiated on this occasion, or **Downward Causation*** x causes any underlying property instance of y to be instantiated on this occasion.

From these premises, we can infer that when M_1 is realized by P_1 on some occasion, m_1, an instance of M_1 on this occasion, will be causally impotent if it is not identical with p_1, an instance of P_1 on this occasion. Suppose that m_1 metaphysically depends on, but is not identical with, p_1 and that m_1 causes m_2, which is an instance of M_2 on this occasion. Also suppose m_2 metaphysically depends on, but is not identical with, p_2, an instance of P_2 on this occasion. According to *Causal Realization** (*Downward Causation**), if m_1 causes m_2, it must also cause p_2. Given *Closure*, p_2 is presumably caused by p_1. Thus, both m_1 and p_1 are purported to be the cause of p_2. However, *Exclusion* and *Non-overdetermination** jointly imply that only one of them is the cause of p_2. To adhere to *Closure*, we should choose p_1 as the cause. In other words, we need to deny that m_1 causes p_2. Given *Causal Realization**, m_1 also does not cause m_2. This shows that every irreducible mental property instances are causally impotent. We may call this argument *the token-causal exclusion argument*.

Not every non-reductive physicalist needs to address the token-causal exclusion argument. There are token-identity theorists who believe that a mental particular is identical with a physical particular even though the mental property and the physical property of which they are instances are distinct from each other (e.g. Davidson 1970; Robb 1997; Heil and Robb 2003). Since token-identity theorists reject the irreducibility of mental particulars, the token-causal exclusion argument poses no threat to their views. Reducible mental particulars are causally efficacious because they share the same causal power with their physical counterparts. Thus, to secure the causal efficacy of the mental, token-identity theorists need to address only the type-causal exclusion argument.

However, many non-reductive physicalists accept not only the irreducibility of mental properties but also the irreducibility of mental particulars formulated as, for example, *Irreducibility**. It might seem to them that the token-identity theory is unpromising because it mischaracterizes the relation between the causal power of a mental property and the causal power of its instances. If m_1 is causally efficacious merely because it is identical with p_1 whose causal power comes from P_1, then M_1 will be explanatorily irrelevant to the causal efficacy of m_1. This runs counter to one of the fundamental intuitions that motivate non-reductive physicalism: mental properties are causally efficacious on their own and we seem to have a lot of first-hand experience with how their causal powers are exercised on everyday occasions. To take this intuition at face value is to assert that m_1 is causally efficacious in light of its being an instance of M_1, rather than its being

identical with p_1. This is why many non-reductive physicalists reject the token-identity theory. I expect causal autonomists to be among them. Thus, to secure the causal efficacy of the mental, causal autonomists should not only reject the type-causal exclusion argument but also the token-causal exclusion argument.

However, the strategy causal autonomists employ to address the type-causal exclusion argument cannot be adopted to cope with the token-causal exclusion argument. To see this, we may begin with constructing an account of token causation parallel to *Property Causation$_2$* Zhong appeals to:

Token Causation$_1$: Property instance x causes property instance y on some occasion if and only if:
(I) If an intervention that makes x present were to occur (while all other relevant variables in the causal graphs are fixed) on this occasion, then y would also be present on this occasion; and
(II) If an intervention that makes x absent were to occur (while all other relevant variables in the causal graphs are fixed) on this occasion, then y would also be absent on this occasion.

Clearly, this is not the only account of token causation a causal autonomist might adopt. Since List and Menzies formulates their account of property causation in terms of possible-world semantics, we may expect them to adopt a similar account of token causation as follows:

Token Causation$_2$: Property instance x causes property instance y if and only if the presence of x makes a difference to the presence of y in the actual world, if and only if it is true in the actual world that:
(I) x is present $\square \rightarrow y$ is present; and
(II) x is absent $\square \rightarrow y$ is absent.

Since our investigation of token-causal relationships usually starts with an observation that one property instance occurs following the occurrence of another, the first conditions in *Token Causation$_1$* and in *Token Causation$_2$* are often trivially met.[9] Thus, in those situations where two property instances, x and y, indeed occur, we can rewrite *Token Causation$_1$* as follows:

Token Causation$_3$: Property instance x causes property instance y on some

---

[9] The first condition in *Token Causation$_2$* is met because List and Menzies adopts Lewis's version of possible-world semantics, according to which the actual world is the closest world to itself.

occasion, if and only if an intervention that makes x absent were to occur (while all other relevant variables in the causal graphs are fixed) on this occasion, then y would also be absent on this occasion.

Similarly, we can also rewrite *Token Causation₂* as follows:

Token Causation₄: Property instance x causes property instance y if and only if the presence of x makes a difference to the presence of y in the actual world, if and only if it is true in the actual world that x is absent $\Box\to y$ is absent.

Note that *Token Causation₄* is exactly the account of token causation adopted by Lewis (1973). *Token Causation₃* and *Token Causation₄* are different in regard to which counterfactual situation is relevant to verify the existence of the causal relationship between x and y. According to *Token Causation₃*, it is the counterfactual situation in which x is absent while every other factor remains in the same state. In contrast, according to *Token Causation₄*, the relevant counterfactual situation is the one that has the same initial background conditions as the actual situation, except x is magically altered to be absent. Although the difference between these two accounts of token causation may sometimes result in different treatments of certain causal scenarios, it makes no difference with regard to the token-causal exclusion argument. Thus, when addressing this argument, causal autonomists can choose either causal account to examine whether certain token causal relationships obtain.

In a situation where M_1 (which is realized by P_1) causes M_2 (which is realized by P_2), we observe the presence of m_1 accompanied by the presence of p_1, and subsequently observe the presence of m_2 accompanied by the presence of p_2. Suppose we are interested in whether m_1 causes p_2 on this occasion, and suppose we choose *Token Causation₃* to verify this. Given this, m_1 causes p_2 on this occasion if and only if:

(5) If an intervention that makes m_1 absent were to occur (while all other relevant variables in the causal graphs are fixed) on this occasion, then p_2 would also be absent on this occasion.

If (5) is true then it entails that m_1 causes p_2.

In order to block the token-causal exclusion argument, causal autonomists like Zhong need to reject *Causal Realization**. In other words, they must show that while m_1 causes m_2 on this occasion, it does not cause p_2. Otherwise, both m_1 and p_1 will be candidate causes of the same effect, which would lead to the causal exclusion of m_1. According to *Token Causation₃*, m_1 causes m_2, if and only if:

(6) If an intervention that makes m_1 absent were to occur (while all other relevant variables in the causal graphs are fixed) on this occasion, then m_2 would also be absent on this occasion.

Therefore, to maintain the causal efficacy of irreducible mental particulars, causal autonomists must prove that when (6) is true, (5) can be false.

However, it is impossible for (5) to be false while (6) is true. Given that p_2 is the *actual* physical realizer of m_2 on this occasion, if m_2 were absent, p_2 would also be absent. That is to say, when (6) is true, (5) must also be true. Thus, according to *Token Causation*$_3$, when m_1 causes m_2, it also causes p_2. Combined with other premises of the token-causal exclusion argument, this claim will lead to the conclusion that m_1 is causally impotent because its actual physical realizer, p_1, does all the causal work on this occasion.

Employing *Token Causation*$_4$ to examine this case also gives us the same conclusion. Given that p_2 is the *actual* physical realizer of m_2 on this occasion, in the closest possible world where m_1 does not occur, m_2 and hence p_2 do not occur, either. Thus, *Token Causation*$_4$ also asserts that if m_1 causes m_2, it also causes p_2. The causal exclusion argument then follows.

In sum, even if causal autonomists can address the type-causal exclusion argument and show how irreducible mental properties can be causally efficacious, their arguments cannot be adopted to prove that irreducible mental particulars are also causally efficacious. This result is unwelcome because it amounts to admitting that while there are autonomous causal laws governing the domain of mental properties, none of these laws are instantiated on any particular occasion. Thus, although mental property instances are not causally impotent, they are causally inert in the sense that it is physical property instances that do all of the *actual* causal work. The failure to address the token-causal exclusion argument indicates that causal autonomism is unsatisfactory

4   Rethinking the Proportionality Requirement

Causal autonomists may agree that their approach cannot successfully reject the token-causal exclusion argument. Nevertheless, they might still hold that it nicely blocks the type-causal exclusion argument. In what follows, I will try to show that this is not the case.

It is the proportionality requirement that serves as the core of the difference-making accounts of property causation favored by causal autonomists. However, it is unclear whether causation requires proportionality or whether our conception of causation is really committed to the proportionality requirement. Some philosophers deny that the proportionality requirement is at the center of our causal intuitions by showing that many of our ordinary causal claims that we view as reliable go against this requirement (e.g. Shapiro and Sober 2012; McDonnell 2017). For example, in the claim "Socrates' drinking hemlock caused him to die," the cause, one's drinking hemlock, is not proportional to the effect,

one's death, because there could be many other means than drinking hemlock which will also lead to one's death. Causal autonomists may reply there is always some common feature between all of the means to one's death such that it is this common feature that is proportional to one's death and hence genuinely causes the latter. However, given that there are a great number of means which can lead to one's death, it is very unlikely, if not impossible, to find any common feature among all of these means except that they all lead to one's death.

For this reason, it seems that *Property Causation*$_1$ can better capture our rationale behind the claim "Socrates' drinking hemlock caused him to die." We regard one's drinking hemlock as a cause of one's death because there is a possible intervention on one's drinking hemlock such that if that intervention were to occur, the state of whether one would die would change when other relevant factors are held fixed.

Perhaps the lesson we should draw is that while some of our causal intuitions are committed to the proportionality requirement, others are not. It is a pragmatic matter for us to decide whether we want to explicitly build the proportionality requirement into our best account of property causation. I do not think this is a hard choice to make. Accepting the proportionality requirement would lead to the abolishing of many ordinary and useful causal attributions, without having any good replacement which could serve the same function. On the other hand, if we abandon the proportionality requirement, we will merely have a larger set of true causal claims to affirm. For example, we will then regard both the redness and scarletness of an object to be the cause of the pigeon's pecking. While this may slightly increase our cognitive tasks, it has no noticeable disadvantages regarding our everyday practices. It is clear that when we have to make a choice between *Property Causation*$_1$ and, for example, *Property Causation*$_2$, it is better, given our role as agents who act and reason based on our causal knowledge, to choose the one that is not committed to the proportionality requirement.

However, there is a third option which can capture the spirit of the proportionality requirement without committing to some of its drawbacks. The spirit of the proportionality requirement is that the cause must always co-vary with the effect. It does not specify which kind of covariance it is. *Property Causation*$_2$ adds that this covariance functions in a deterministic manner: The presence (or absence) of X is necessarily associated with the presence (or absence) of Y. Given that we have rejected *Property Causation*$_2$, we should no longer stick to the deterministic understanding of covariance. Once the latter requirement is loosened, we can have an alternative account of property causation as follows.

Property Causation$_4$: Property X causes property Y if and only if:
(I) If an intervention that sets $X = x_p$ were to occur (while all other relevant variables in the causal graphs are fixed), then the probability that $Y = y_p$ would rise; and

(II) If an intervention that sets $X = x_a$ were to occur (while all other relevant variables in the causal graphs are fixed), then the probability that $Y = y_a$ would rise.

Now the covariance is not between the presence (or absence) of X and the presence (or absence) of Y, but between the presence (or absence) of X and the probability of the presence (or absence) of Y. This modification enables *Property Causation₄* to better handle our causal attributions compared with either *Property Causation₂* or *Property Causation₃*.

However, once causal autonomists move from the accounts of property causation, they favor to either *Property Causation₁* or *Property Causation₄*, they become unable to reject *Causal Realization* because they cannot show that when M_1 causes M_2, M_1 does not cause P_2. According to *Property Causation₁*, M_1 causes P_2, if and only if:

(7) There is a possible intervention on M_1 such that if that intervention were to occur, the value of P_2 would change when other relevant factors are held fixed. Similarly, M_1 causes M_2, if and only if:
(8) There is a possible intervention on M_1 such that if that intervention were to occur, the value of M_2 would change when other relevant factors are held fixed.

To reject *Causal Realization*, causal autonomists must show that when (8) is true, (7) can be false. But this cannot follow from *Property Causation₁*. Because M_2 can be realized by P_2, when there is a possible intervention on M_1 that leads to a change in the value of M_2, there is also a possibility that this intervention on M_1 also leads to a change in the value of P_2. Thus, when (8) is true, (7) must also be true.

Property Causation₄ does not have the result desired by causal autonomists, either. According to *Property Causation₄*, M_1 causes P_2, if and only if:

(9) If an intervention that makes M_1 present were to occur (while all other relevant variables in the causal graphs are fixed), then the probability that P_2 is present would rise; and
(10) If an intervention that makes M_1 absent were to occur (while all other relevant variables in the causal graphs are fixed), then the probability that P_2 is absent would rise.
Similarly, M_1 causes M_2, if and only if:
(11) If an intervention that makes M_1 present were to occur (while all other relevant variables in the causal graphs are fixed), then the probability that M_2 is present would rise; and
(12) If an intervention that makes M_1 absent were to occur (while all other

relevant variables in the causal graphs are fixed), then the probability that M_2 is absent would rise.

To reject *Causal Realization*, causal autonomists must show that when (11) and (12) are true, (9) and (10) can be false.

Suppose the conditional probability that M_2 is realized by P_2 given that M_2 occurs is n ($0 < n < 1$ because P_2 is only one among many physical realizers of M_2). Further, suppose M_1 causes M_2 such that the conditional probability that M_2 is present given that M_1 is present is l, while the conditional probability that M_2 is present given that M_1 is absent is d. Because (11) and (12) are true, $0 \leq d < l \leq 1$. So when M_1 is present, there is a chance of ln that P_2 is present (i.e. $\Pr(P_2|M_1) = ln$). Accordingly, the probability that P_2 is absent given that M_1 is present is $1 - ln$ (i.e. $\Pr(\sim P_2|M_1) = 1 - ln$). When M_1 is absent, there is only a chance of d that M_2 is present. Thus, the probability that P_2 is present given that M_1 is absent is dn (i.e. $\Pr(P_2|\sim M_1) = dn$). Accordingly, the probability that P_2 is absent given that M_1 is absent is $1 - dn$ (i.e. $\Pr(\sim P_2|\sim M_1) = 1 - dn$). Clearly, $\Pr(P_2|M_1) > \Pr(P_2|\sim M_1)$ because $d < l$. So (9) is true. Similarly, it is clear that $\Pr(\sim P_2|\sim M_1) > \Pr(\sim P_2|M_1)$. So (10) is also true.

Thus, according to *Property Causation₄*, when M_1 causes M_2, it also causes P_2. That is to say, causal autonomists are unable to reject *Causal Realization* by appealing to *Property Causation₄*. Given that none of the accounts of property causation favored by causal autonomists can better support their claims, the overall strategy adopted by causal autonomists cannot successfully address the type-causal exclusion argument.

5   Conclusion

I have examined the causal autonomists' strategy and tried to show that it cannot successfully address the causal exclusion argument. For their strategy to secure the causal efficacy of the mental, it needs to block the type-causal exclusion argument which denies the causal efficacy of irreducible mental properties, as well as the token-causal exclusion argument which denies the causal efficacy of irreducible mental particulars. However, their arguments fail to meet both requirements. In order to defend their view, non-reductive physicalists should either seek some other grounds for causal autonomism or deny that it can serve as a promising response to the causal exclusion argument.

References

Campbell, J. 2010a. "Control Variables and Mental Causation." *Proceedings of the Aristotelian*

Society 110.1: 15–30.

Campbell, J. 2010b. "Independence of Variables in Mental Causation." *Philosophical Issues* 20.1: 64–79.

Davidson, D. 1970. "Mental Events," in *Essays on Actions and Events*, edited by L. Foster & J. W. Swanson, 207–24. Oxford: Clarendon Press.

Fodor, J. 1974. "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)." *Synthese* 28.2: 97–115.

Heil, J., and David R. 2003. "Mental Properties." *American Philosophical Quarterly* 40.3: 175–96.

Kim, J. 1976. "Events as Property Exemplifications," in *Action Theory*, edited by M. Brand & D. Walton, 310–26. Dordrecht: Reidel.

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, NJ: Princeton University Press.

Lewis, D. 1973. "Causation." *Journal of Philosophy* 70.17: 556–67.

List, C., and Peter Menzies. 2009. "Nonreductive Physicalism and the Limits of the Exclusion Principle." *Journal of Philosophy* 106.9: 475–502.

McDonnell, N. 2017. "Causal Exclusion and the Limits of Proportionality." *Philosophical Studies* 174.6: 1459–74.

Putnam, H. 1967. "Psychological Predicates", in *Art, Mind, and Religion*, edited by W. H. Capitan & D. D. Merrill, 37–48. University of Pittsburgh Press.

Robb, D. 1997. "The Properties of Mental Causation." *Philosophical Quarterly* 47.187: 178–94.

Shapiro, L., and Elliott S. 2007. "Epiphenomenalism—The Do's and Don'ts," in *Thinking about Causes: From Greek Philosophy to Modern Physics*, edited by Peter Machamer & Gereon Wolters, 235–64. Pittsburgh: University of Pittsburgh Press.

Shapiro, L., and Elliot S. 2012. "Against Proportionality." *Analysis* 72.1: 89–93.

Woodward, J. 2008. "Mental Causation and Neural Mechanisms," in *Being Reduced: New Essays on Reduction, Explanation, and Causation*, edited by Jakob Hohwy & Jesper Kallestrup, 218–62. Oxford: Oxford University Press.

Woodward, J. 2010. "Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation." *Biology and Philosophy* 25.3: 287–318.

Woodward, J. 2015. "Interventionism and Causal Exclusion." *Philosophy and Phenomenological Research* 91.2: 303–47.

Yablo, S. 1992. "Mental Causation." *Philosophical Review* 101.2: 245–80.

Zhong, L. 2014. "Sophisticated Exclusion and Sophisticated Causation." *Journal of Philosophy* 111.7: 341–60.