

Mental Causation for Standard Dualists

Bram Vaassen

September 12, 2023

Forthcoming in *Australasian Journal of Philosophy*
Penultimate Draft, please cite published version

Abstract

The standard objection to dualist theories is that they have a hard time accounting for the seemingly obvious fact that these mental phenomena can cause our behaviour. On the plausible assumption that all our behaviour is physically necessitated by entirely physical phenomena, there appears to be no room for dualist mental causation. Some argue that dualists can address this problem by making minimal adjustments in their ontology. I argue that no such adjustments are required. Given current trends in philosophy of causation, mental phenomena stand a good chance of being causes, even in standard dualist ontologies.

Keywords: Mental Causation, Dualism, Consciousness, Causal Exclusion, Higher-Level Causation

1 Introduction

Dualists maintain that phenomenally conscious mental phenomena are metaphysically distinct from all physical phenomena (e.g., Chalmers, 1996). The standard objection to dualist theories is that they have a hard time accounting for the seemingly obvious fact that these mental phenomena can cause our behaviour. On the plausible assumption that all our behaviour is physically necessitated by entirely physical phenomena, there appears to be no room for dualist mental causation (e.g., Papineau, 2002; Bennett, 2008; Chalmers, 2013; Goff, 2017). For a long time, this objection reassured many of us who are inclined towards physicalism that dualism will always come at a nearly unbearable cost.

In recent years, this reassurance has been challenged. Several authors have argued that minimal adjustments to the dualist ontology could deliver mental causation after all (e.g., Bealer, 2007; Lowe, 2008; Lycan, 2009; Gibb, 2015; Kroedel, 2015, 2020; Won, 2021; Khawaja, 2022).¹ I will argue that the situation is even worse. Recent results from the causation literature point the way towards a strategy the dualist can use to save her theory without even having to make minimal adjustments to her ontology.

I will make my case as follows. First, I characterize a standard dualist position and explain why it is widely assumed not to allow for mental causation (§2). In §3, I present the interventionist framework and discuss a familiar strategy for capturing the right amount of higher-level causation defended by Yablo (1992); Papineau (2013); Blanchard (2020); Woodward (2020); Zhong (2020b,a), and myself Vaassen (2022). In §4, I argue that a similar strategy can be used to provide an interventionist account that allows for dualist mental causation. I also address the objection from causal exclusion and the objection that the resulting account is too permissive. With these objections out of the way, I conclude that the standard objection against standard dualism fails.

¹See also List and Stoljar (2017), who do not provide an account of dualist mental causation (p. 105), but instead argue against so-called ‘exclusion arguments’ (See also §4.3).

2 Standard Dualism

Dualists are concerned with the ontological status of *phenomenally conscious* mental phenomena. Phenomenally conscious mental phenomena are those mental phenomena that constitute *what it is like to be* for someone. For example, feeling pain and smelling sour milk constitute what it is like for someone to be. According to the dualist, such phenomenally conscious mental phenomena are *metaphysically distinct* from physical phenomena, such as leptons and quarks having a certain spin, mass and location, in the following sense:²

Metaphysical Distinctness For any two phenomena *A* and *B*, *A* and *B* are *metaphysically distinct* if and only if it is metaphysically possible for *A* to occur without *B* occurring, *and vice versa*.

That is to say, for any physical phenomenon and any phenomenally conscious mental phenomenon, there is possible world where one occurs but the other does not, and *vice versa*. For example, there is possible world that is physically identical to ours, but lacks phenomenally conscious mental phenomena altogether.³ Or so claims the dualist (e.g., Chalmers, 1996, p. 94–99).

Three further commitments relevant to mental causation typically feature in dualist ontologies. First, dualists typically maintain that mental phenomena are unique in their metaphysical distinctness from physical phenomena. Plausibly, non-mental phenomena, such as hurricanes, banking crises, and milk going sour, are metaphysically necessitated by physical phenomena. That is to say, for any actual higher-level phenomenon *H*, there is some actual physical phenomenon *P*, such that any world containing *P* also contains *H*. I will call the physical phenomenon that metaphysically necessitates a higher-level phenomenon the ‘physical metaphysical base’ of that phenomenon. The dualist must deny that

²Cf., List and Stoljar (2017, p. 98).

³For simplicity, I will drop the qualifier ‘phenomenally conscious’ when talking of phenomenally conscious mental phenomena.

mental phenomena have physical metaphysical bases, but she allows for non-mental phenomena to have them.

Second, dualists typically maintain that mental phenomena are *nominally* necessitated by physical phenomena (e.g., Chalmers, 1996, p. 126). To explain the pervasive correlations between mental phenomena and physical phenomena, dualists posit psychophysical laws of nature that are modally on a par with the fundamental laws of physics in that they are *fundamental*, but *contingent*. They are fundamental in that they are (i) exceptionless and (ii) not explained by any further laws or facts. They are contingent in that they do not hold in all possible worlds: there are worlds where physical phenomena fail to give rise to the relevant mental phenomena. I will refer to such contingent, fundamental laws as '*nomic laws*'. The dualist thus typically allows for mental phenomena to have physical *nomic* bases, whilst denying that they have physical *metaphysical* bases.

Finally, the dualist typically accepts PHYSICAL COMPLETENESS:

Physical Completeness For any actual physical phenomenon P and any time t , there is a purely physical phenomenon that occurs at t and physically necessitates the occurrence of P .⁴

Where A physically necessitates B if and only if all physically possible worlds that contain A also contain B and by 'physically possible world', I mean any possible world in which the same fundamental laws of physics as in our world hold. As PHYSICAL COMPLETENESS is widely treated as a necessary ingredient for scientifically respectable ontologies, it is in the dualist's interest to accept it.

We now arrive at the following position:

Standard Dualism For any mental phenomenon M and any physical phenomenon P , M and P are metaphysically distinct and:

⁴We assume determinism for simplicity. Translations to indeterministic-friendly formulations are straightforward (cf., Bennett, 2008, fn. 3).

- i All non-fundamental, non-mental phenomena have physical metaphysical bases
- ii All mental phenomena have physical nomic bases
- iii PHYSICAL COMPLETENESS

There are well-known arguments to the effect that STANDARD DUALISM cannot allow for mental phenomena to cause our behaviour (cf., §4.3), but we don't need to look at them in detail in order to appreciate their pull. Given (i) and (ii), all our bodily movements are physically necessitated by prior physical phenomena. This seems to leave no room for dualist mental phenomena to cause our bodily movements. Dualists and non-dualists alike tend to agree that STANDARD DUALISM cannot allow for mental causation (e.g., Chalmers, 1996, 2013; Papineau, 2002; Bennett, 2008; Goff, 2017).

Consequently, philosophers who favour dualist mental causation typically propose adjustments to STANDARD DUALISM. For example, Bealer (2007); Lowe (2008); Gibb (2015); Won (2021) and Kroedel (2015, 2020) provide accounts of dualist mental causation by making minimal concessions on the completeness of the physical or the modal strength of psychophysical laws.⁵ I submit that no such concessions are required. Given current trends in philosophy of causation, there is a plausible account of causation that allows for mental causes, even if STANDARD DUALISM is true.

It will help my exposition to start from an existing account of causation. The arguments I propose would fit with a number of respectable accounts of causation, including the statistical account by Papineau (2022) and the thermodynamical accounts by Loewer (2007a), Albert (2015), and Ismael (2016). I will restrict my focus to the interventionist accounts defended by Woodward (2008,

⁵Some of these proposals will also require further empirical commitments. For example, Gibb's proposal requires that all mental causation happens via a specific double preventer structure (Gibb, 2013, § 4). In earlier work, Lowe also proposed accounts with particular empirical commitments (1996; 1999). See Vaassen (2019, Ch. 8) for discussion of the Lowe's proposals. The account provided here will require no more empirical commitments than are standard in physicalist accounts of mental causation (cf., §4).

2015); Campbell (2010); Raatikainen (2010); Zhong (2020a,b), and others. Given their prevalence in the literature on both higher-level causation and mental causation, such accounts provide a natural starting point. My strategy is to argue that, in as far as these accounts can be made to deliver higher-level causation and mental causation in physicalist ontologies, they can be made to deliver mental causation in dualist ontologies as well.

Before continuing, an important caveat needs to be made. While starting from the interventionist account helps to ground my arguments in the literature, it also limits their scope. Interventionist accounts may be popular, they are not undisputed. These accounts belong to a family of ‘lightweight’ theories of causation that do not require causes to produce or necessitate their effects, but characterize causation in terms of *dependence* instead. One important worry is that such lightweight approaches fail to capture what is actually at issue in mental causation debates (e.g. Kim, 2005, p. 17–18). If mental causation indeed requires some causal *oomph* that is lacking in interventionist accounts and their lightweight cousins, then the argument that follows does not provide the dualist with mental causation. I don’t think this caveat should bother the dualist too much. Given that physicalists typically rely on lightweight theories to account for mental causation as well, the argument still carries considerable dialectical weight. In fact, I think it is likely that, if dependence theories fail, both physicalists and dualists are in trouble. But let us not get ahead of ourselves. We can start with a look at the interventionist approach, and reserve concerns about its lack of *oomph* for §4.4.

3 Interventionism

The idea driving interventionism is commonsensical. Causal interactions with causes correlate with changes in their effects. Consequently, we can, at least in principle, manipulate effects by manipulating their causes. For example, causing

someone to smoke less correlates with a decrease in their risk of incurring lung cancer. We can therefore manipulate lung cancer risk by manipulating smoking behaviour. By contrast, causal interactions with phenomena that do not cause lung cancer, such as the colour of my shoes, tend not to correlate with changes in that effect. While most accounts acknowledge this connection between causation and manipulability, the interventionist takes it a step further and proposes to characterize causation in terms of correlations that are due to manipulations.⁶

Of course, sometimes manipulations of non-causes of an effect *do* correlate with changes in that effect. This is the case when the manipulation of the non-cause also interacts with a cause of the target effect. For example, manipulating tar-stained fingers by causally interacting with smoking behaviour will correlate with lung cancer risk, even though tar-stained fingers do not cause lung cancer. Any account of causation needs a principled way of distinguishing such confounded correlations from causal correlations.

The interventionist can spell out the relevant difference in terms of manipulability as well (cf., Woodward, 2014, p. 710). The correlation between manipulations of tar-stained fingers that go via smoking behaviour and cancer risk are due to an idiosyncrasy of those manipulations: they also affect smoking behaviour. Manipulations of tar-stained fingers that do not share this idiosyncrasy should not be expected to correlate with changes in cancer risk. By contrast, manipulations of smoking behaviour will correlate with changes in cancer risk even if they do not surreptitiously affect an independent cause of cancer risk. Even idealized manipulations that manage to solely affect the smoking behaviour and keep everything but the downstream effects of the change in smoking behaviour unaffected can be expected to correlate with changes in cancer risk. Given that no idiosyncrasies are required for manipulations of smoking behaviour to correlate with cancer risk, lung cancer risk is more reliably manipulable through smoking behaviour than through tar-stained fingers.

⁶Cf., Franklin-Hall (2016, p. 556) and Kaiserman (2020, p. 920).

Let us call such idealized manipulations that are riddled of disturbing idiosyncrasies ‘interventions’. We can then present a rough sketch of the interventionist picture as follows (cf. Woodward, 2008; Zhong, 2014):⁷

INTERVENTIONISM I

Causation I *X* causes *Y* if and only if there is a possible intervention on *X* relative to *Y* that would be followed by a change in the value of *Y*.

Intervention I A manipulation of *X* counts as an intervention on *X* relative to *Y* if and only if it causally changes the value of *X* whilst holding all other causes of *Y* fixed at some value.

Following interventionist custom, these definitions are formulated in terms of variables such as *X* and *Y*, where variables represent properties or quantities that can take at least two values, such as *mass* or *cigarettes smoked per day* (Woodward, 2008, p. 222). For simplicity, we will mostly focus on two-valued variables that represent a phenomenon occurring (value = {1}) or not occurring (value = {0}).⁸ According to INTERVENTIONISM I, Sam being a smoker is a cause of his lung cancer if and only if some possible manipulation of his smoking behaviour that left all other causes of lung cancer unaffected would be followed by a change in his incurring lung cancer (cf., Figure 1).

While INTERVENTIONISM I is only a rough sketch of interventionist accounts, it demonstrates an important feature of these accounts: the ‘holding fixed’-requirement in *Intervention I* stands in the way of dualist mental causation. Consider a simple case of my being in pain and subsequently wincing in a dualist ontology. My pain *Pa* is nomically necessitated by a physical phenomenon *Ph*,

⁷I follow Zhong (2014), Woodward (2020), and others by setting aside complications involving causal intermediaries.

⁸This simplification is common in mental causation debates (Woodward, 2008, p. 222; Zhong, 2014, p. 344).

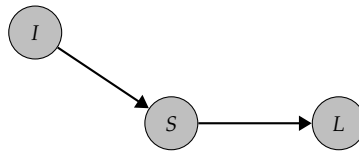


Figure 1: The nodes represent the causal relata and thin single arrows represent causal relations. Interventions (I) on being a smoker (S) correlate with changes in lung cancer (L). Thus, S is a cause of L .

which is a part of the physically sufficient condition leading up to my wincing W_i . Given PHYSICAL COMPLETENESS, any intervention on Pa which 'holds fixed' all other phenomena that cause W_i will not correlate with changes in my wincing.

It thus appears that there is a firm barrier against dualist mental causation. The 'holding fixed'-requirement plays a crucial role in INTERVENTIONISM I and prevents dualist mental phenomena from being causes. However, recent efforts to capture *higher-level* causations demonstrate that exceptions to the requirement can be made if one treads with appropriate care. The strategy employed in these efforts can be employed by the dualist as well.

3.1 Higher-level causation

Causes like hurricanes, banking crises and infections are metaphysically necessitated by phenomena that are individuated in terms of leptons and quarks having a certain location and spin, but they are not themselves individuated in those terms.⁹ Any respectable account of causation must capture such *higher-level* causation. In the following three subsections, I summarize a popular strategy for capturing the right amount of higher-level causation in an interventionist framework. Afterwards, I argue that this strategy provides dualists with the tools to build an account of mental causation.

⁹This is for standard reasons of multiple realizability; for any banking crisis, it is the case that the very same banking crisis could have occurred whilst its physical metaphysical base contained some leptons or quarks more or less.

We can begin with a challenge for INTERVENTIONISM I. Some argue that its ‘holding fixed’-requirement threatens higher-level causation (e.g., Baumgartner, 2010). Consider a plausible higher-level causation claim like ‘the 2008 banking crisis caused a rise in unemployment’. The 2008 banking crisis has a physical metaphysical base. Unless we can exclude that base from being a cause of the subsequent rise in unemployment, *Intervention I* demands that interventions on the 2008 banking crisis relative to unemployment numbers should hold this base fixed. However, manipulating the banking crisis without making a change in its physical base is metaphysically impossible. Consequently, there are no possible interventions on the banking crisis relative to the unemployment numbers — let alone interventions that are followed by changes in those numbers — and the banking crisis cannot be a cause of the rise in unemployment according to INTERVENTIONISM I.

In response, interventionists note that applying the ‘holding fixed’-requirement to metaphysical bases defeats its original purpose (e.g., Woodward, 2014, 2015). We introduced this requirement to isolate correlations that are available for reliable manipulation from those that are not. To do so, we demanded that causation-revealing manipulations are free of any deceptive idiosyncrasies. However, affecting the metaphysical base of a higher-level phenomenon is not just an idiosyncrasy of any specific manipulation; it is a metaphysically necessary condition for any manipulation of a higher-level phenomenon. Consequently, there is no real risk of manipulations of a higher-level cause failing to bring about a change in the target effect because they failed to affect the metaphysical base of the higher-level cause — whereas there is often a real risk of manipulating a phenomenon without manipulating phenomena that are only contingently related to it. While the interventionist’s goal to characterize causation in terms of reliable manipulability motivates holding fixed metaphysically distinct variables when intervening, it does not appear to motivate holding fixed variables that stand in metaphysical necessitation relations.

To allow for higher-level causation, we can add an exemption clause for variables whose values metaphysically necessitate, or are metaphysically necessitated by, the values of the purported cause or the target effect:

INTERVENTIONISM II

Causation II X causes Y if and only if there is a possible intervention on X relative to Y that would be followed by a change in the value of Y .

Intervention II A manipulation of X counts as an intervention on X relative to Y if and only if it causally changes the value of X whilst holding all other causes of Y fixed at some value, *except for those that stand in a metaphysical necessitation relation to the cause or effect variable.*

As it is no longer required to hold its metaphysical base fixed whilst intervening on the 2008 banking crisis, there is no reason to believe that interventionist accounts exclude it from having caused the subsequent rise in unemployment (cf., Figure 2).

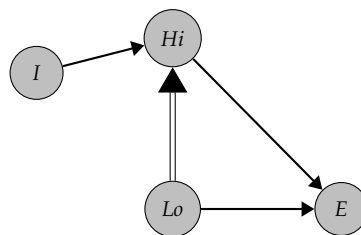


Figure 2: The double arrow represents metaphysical necessitation. Higher-level phenomenon Hi is metaphysically necessitated by lower-level phenomenon Lo ; both cause E , because interventions like I correlate with changes in E .

With the adjusted 'holding fixed'-requirement, interventionism allows for higher-level causation. The problem is that it allows for too much.

3.2 Spurious Higher-Level Correlations

INTERVENTIONISM II is too permissive. It predicts that higher-level causes always inherit the causal powers of their underlying phenomena. But this is not the case.

Consider the ‘p’ that appears on my computer screen when I press the ‘p’ key. The nature of the electron cloud in the electrical wire connecting my keyboard to my screen is such that the wire conducts electricity, and it is thereby that the ‘p’ appears. The nature of the electron cloud also metaphysically necessitates the opacity of this wire, but the opacity of the wire did not cause the ‘p’ to appear on my screen. The conductivity and the opacity are both metaphysically necessitated by the nature of the electron cloud. The former inherits the causal role relative to the ‘p’ appearing, but the latter does not.¹⁰

INTERVENTIONISM II fails to distinguish the causal roles of the conductivity and the opacity of the wire. There are possible interventions on the opacity of the wire that would be followed by the ‘p’ not appearing on my screen rather than appearing on my screen. For example, an intervention that changed the electrical wire with a transparent plastic wire would have that result. Of course, such an intervention would also affect the nature of the electron cloud in the wire, but that is allowed by *Intervention II*. These definitions spuriously count the opacity of the wire as a cause of the ‘p’ appearing.

Similar counterexamples are easy to generate and can readily be found in the literature. Consider Alice, a pigeon trained to peck exclusively at scarlet objects (List and Menzies, 2009, p. 494).¹¹ She is presented with an scarlet pebble, which causes her to peck. The pebble being scarlet also metaphysically necessitates its being red. Any intervention on the pebble being red would result in her not pecking. Nevertheless, we would not want to say that Alice pecked because the pebble was red. After all, she would not have pecked if

¹⁰Cf., Jackson and Pettit (1990, p. 204). See also Walter (2010).

¹¹See also Yablo (1992) and Zhong (2020b, p. 34).

the pebble had been red, but not scarlet. Or consider physical phenomena at the lowest, most fundamental level, such as leptons and quarks having their exact spin, charge and location. Such phenomena are sensitive to even the minutest changes at enormous distances (cf., Gallow, 2015, §1.3). The slightest intervention on the location and spin of the fundamental particles making up Napoleon's big blue hat during the battle of Waterloo, would have made for changes in the location and spin of the fundamental particles making up the eggs I had for breakfast. And yet, we would not want to say that Napoleon wearing a big blue hat during the battle of Waterloo caused the precise physical constitution of my eggs being runny. In these cases as well, interventions on a causally irrelevant higher-level phenomenon would be followed by a change in the target effect. INTERVENTIONISM II systematically counts such spurious higher-level correlations as causal.

Adjusting the 'holding fixed'-requirement made the account too permissive. We need to add a further requirement to pick up the slack.

3.3 Robustness

There is a broadly accepted strategy for dealing with spurious higher-level correlation. The key insight driving such strategies is that the correlation between higher-level causes and their effects is insensitive to changes in the lower-level variables underlying the cause in a way that spurious higher-level correlation is not.¹² Adding an insensitivity requirement on higher-level correlations will help to capture just the right amount of higher-level causation.

Here is one way to capture the difference between higher-level causation and spurious correlations. First, let us define 'screening off' as follows:

¹² This insight plays a central role in so-called 'proportionality'-requirements on causation Yablo (1992); List and Menzies (2009); Raatikainen (2010); Papineau (2013); Zhong (2020a,b); Blanchard (2020) and Woodward (2020). As I have argued elsewhere, ROBUSTNESS can be imposed without imposing proportionality, and doing so avoids the challenges faced by proportionality accounts (Vaassen, 2022).

Screening Off For any three non-identical variables A , B and C , A *screens off* the correlation between B and C if and only if, all else being equal, for every value x of A , if A is held fixed at x , then changes in the value of B do not correlate with changes in the value of C .

The screening off relation provides a reliable guide for causation (cf., Papineau, forthcoming). For example, the variable representing scarletness screens off the correlation between the variables representing different shades of scarlet and Alice's pecking, because, with all else equal, once one holds scarletness fixed at either {1} or {0}, changes in the specific shade of scarlet will no longer correlate with changes in Alice's pecking. Similarly, smoking screens off the correlation between tar-stained teeth and lung cancer, because, with all else equal, once one holds smoking fixed at a particular value, changes in the tar-stainedness of teeth no longer correlate with lung cancer. By contrast, redness fails to screen off the correlation between changes in scarletness and changes in pecking. The same holds for the opacity of the wire, Napoleon's hat and their respective underlying phenomena and target effects. It also holds for tar-stained finger tips, smoking behaviour and lung cancer risk: if one holds fixed the tar-stainedness of finger tips, there will still be a correlation between changes in smoking behaviour and lung cancer risk.

Focusing on the spurious higher-level correlation cases, we can now formulate a further requirement on causation. Call it 'robustness':

Robustness For any two variables A and B , the correlation between values of A and values of B is robust, if and only if A screens off the correlations between its base variables and B .¹³

¹³ I take 'base' to be neutral between 'metaphysical base' and 'nomic base'. For this account to work, it is also important that one is not allowed to gerrymander a base variable in function of the effect so as to make the screening off patterns misleading (cf. Franklin-Hall, 2016). As shown by Woodward (2018), and Blanchard (2020), this restriction can be motivated within an interventionist framework. See also fn. 16.

Intuitively, the robustness requirement distinguishes the good cases from the bad. If Alice pecks at all scarlet objects, we can expect that the pebble being scarlet screens of the correlation between changes in its bases, but we should not expect the same to hold for the pebble's redness. Consequently, robustness-like requirements on higher-level causation are commonplace in the literature on causation, and it appears that we have found a way to pick up the slack left by relaxing the 'holding fixed'-requirement.

However, there is a final difficulty to be dealt with. Robustness-like requirements on causation are too demanding when imposed without restriction. In the actual world, whether or not a higher-level phenomenon is followed by its target effect is always somewhat dependent on its physical metaphysical base. This is because phenomena at most fundamental physical level can be 'thermodynamically abnormal': they can be realized such that their future behaviour is erratic at higher levels. For example, the physical metaphysical base of a pebble being scarlet might be realized such that it will suddenly eject a particle with an acceleration powerful enough pierce Alice's skull, killing her on the spot. Certainly, the correlation between being scarlet and Alice's pecking is not insensitive to such thermodynamically abnormal changes in its physical base. As an empirical matter of fact, these observations generalize and all higher-level phenomena have physically possible metaphysical bases with erratic higher-level futures (cf., Albert, 2015). If we insist on robustness across *all* physically possible scenarios, there will be no higher-level causation and we are back at square one.

A natural solution is to restrict the set of scenarios we use to assess robustness (cf., Woodward, 2018, §5). While thermodynamically abnormal realizations are not physically impossible, they are improbable enough that we can safely rely on phenomena never having thermodynamically abnormal physical metaphysical bases in the actual world. For example, the fact that the pebble being scarlet screens off the correlation between its normal physical bases still allows us to reliably manipulate Alice's pecking with scarlet objects. If a robustness-like

approach is to be successful, this degree of robustness will have to allow for higher-level causation as well. We can say that those correlations that are robust across thermodynamically normal changes in the base variables are ‘sufficiently robust’.

The resulting picture might not be as pristine as we would have liked, but it has found powerful defenses in Loewer (2007a), Albert (2015), and Ismael (2016). Similar restrictions on causally relevant scenarios are also employed to address other puzzles about causation, such as causation by absence (McGrath, 2005), the direction of causation (Albert, 2015), and proportionality (Touborg, 2022). Moreover, a sufficient robustness requirement would fit the interventionist framework particularly well. Robustness tracks a correlation’s insensitivity to changes in the bases of the first phenomenon, and thereby tracks their availability for reliable manipulation and prediction: we can reliably predict and manipulate Alice’s pecking behaviour by focusing on scarletness, but not by focusing on redness.

We can now present the following account:

INTERVENTIONISM III

Causation III *X* causes *Y* if and only if (i) there is a possible intervention on *X* relative to *Y* that would be followed by a change in the value of *Y* and (ii) the *X*-*Y* correlation is sufficiently robust.

Intervention III A manipulation of *X* counts as an intervention on *X* relative to *Y* if and only if it causally changes the value of *X* whilst holding all other causes of *Y* fixed at some value, *except for those that stand in a metaphysical necessitation relation to the cause or effect variable*.

The robustness requirement picks up the slack from the relaxed ‘holding fixed’-

requirement, thus delivering an account that allows for just the right amount of higher-level causation.

At first blush, none of this helps the dualist. The ‘holding fixed’-requirement still applies to *nomic* bases and still stands in the way of dualist mental causation. However, the discussion on higher-level causation should have made us wonder whether there is a principled reason for excluding dualist mental phenomena from being causes. Higher-level phenomena are ascribed causal status in virtue of exhibiting sufficiently robust patterns of correlation with target effects. Why couldn’t dualist mental phenomena exhibit such patterns? Moreover, we just reviewed a concrete strategy for relaxing the ‘holding fixed’-requirement without allowing for too much causation: just let the robustness requirement pick up the slack. As physicalists, we better hope that the dualist cannot re-apply this strategy to nomic bases without creating serious problems. In the next section, I argue that, surprisingly, no such problems arise.

4 Dualist Mental Causation

In effect, exempting nomic bases from the holding-fixed requirement amounts to restricting the kinds of scenarios that are relevant when assessing causal claims. When, for example, asking whether my pain causes my wincing, we would not consider the nomically impossible world where my pain has changed without its nomic base changing. Instead, we would consider the nomically possible world where my pain is removed and its nomic base is removed as well. Do we have any principled reason for resisting such a restriction?

First off, note that INTERVENTIONISM III already relies on a similar restriction when evaluating robustness. To rescue higher-level causation, we focused on worlds where the fundamental laws of nature were respected and histories develop in a thermodynamically normal way. By doing so, we allowed the fundamental laws of nature and thermodynamical abnormality to restrict the

kinds of worlds that are causally relevant.

Once we are made aware of such thermodynamically abnormal possibilities, it becomes clear that we require similar restrictions at the intervention stage as well. For example, consider the exact physical realization of any asteroid an hour before I have breakfast. There is some intervention on that phenomenon that correlates with changes in my having breakfast. Changing the physical realization into a thermodynamically abnormal or physically impossible realization that will make the asteroid crash into the earth right where I live would stop me from having breakfast. Similar observations hold for most physical realizations in my backward-looking light cone, and, if we allow for nomically impossible scenarios, beyond. Nonetheless, the physical realizations of distant asteroids aren't causes of my eating breakfast. The robustness requirement will not be of any help here, as the phenomena under consideration are stipulated to take place at the fundamental level. The variables representing those phenomena thus have no base variables, and the relevant correlations are trivially robust.

Again, the problem arises because we consider wildly erratic worlds when testing for causation. And again, the natural response is to exclude such worlds by focusing on nomically possible, thermodynamically normal worlds, as proposed by Loewer (2007a), Albert (2015), and Ismael (2016). As before, this strategy is in the spirit of the interventionist approach, and derives support from its usefulness for addressing other puzzles.¹⁴

If we implement these restrictions in our account of causation, we arrive at:

INTERVENTIONISM IV

Causation IV *X* causes *Y* if and only if (i) there is a possible intervention on *X* relative to *Y* that would be followed

¹⁴See also Goodman (2015); Dorr (2016) and Dorst (2022) for restrictions to nomically possible worlds specifically. For contrast, see Woodward (2003, §3.5), who explicitly proposes that interventions need not be nomically possible. It is unclear how Woodward intends to avoid counterexamples like the asteroid case.

by a change in the value of Y and (ii) the X - Y correlation is sufficiently robust.

Intervention IV A manipulation of X counts as an intervention on X relative to Y if and only if (i) it causally changes the value of X whilst holding all other causes of Y fixed at some value, *except for those that stand in a metaphysical necessitation relation or a synchronic nomic necessitation relation to the cause or effect variable* and (ii) does not take the variable to a thermodynamically abnormal realization.

Interventions on my pain are now allowed to affect their nomic physical base as well. Plausibly, such interventions would correlate with changes in my wincing (Figure 3). The remaining question is whether that correlation is robust. It seems plausible to me that they are.

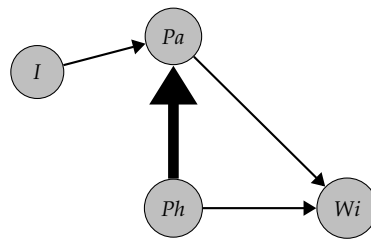


Figure 3: The thick full arrow represents the nomic necessitation relation between Pain (Pa) and (Ph); both of which cause my wincing (Wi)

For simplicity, we can assume that Pa , Ph and Wi can take only two values: $\{1\}$ and $\{0\}$. If Pa takes $\{0\}$, then Ph must take $\{0\}$ as well in order to respect the restriction to nomically possible worlds. If Pa takes value $\{1\}$, then Ph can take either $\{0\}$ or $\{1\}$. Or at least, that is the case if pain is multiply realizably by physical bases.¹⁵ On the further assumption that normal occurrences of pain

¹⁵If this assumption fails, the Pa - Wi correlation is trivially robust.

are followed by wincing independently of what their actual physical base is, it follows that the Pa screens off the correlation between Ph and Wi . Although neither assumption is trivial, both are reasonably well-accepted and play a crucial role in physicalist accounts of mental causation by Yablo (1992); Woodward (2008, 2015); Campbell (2010); Raatikainen (2010) and Zhong (2014, 2020a,b). I see no reason why the dualist should not be allowed to assume the same correlation patterns.¹⁶ The expanded exception to the ‘holding fixed’-requirement opened the door to dualist mental causation. The robustness requirement does not appear to close that door. It is now up to us, the physicalists, to point out what is wrong with INTERVENTIONISM IV. I see four possible objections to INTERVENTIONISM IV. Unfortunately, none of them strike me as particularly powerful.

4.1 The physical-psycho-physical split

In principle, we could argue that *physics* and *thermodynamics* should constrain the causally relevant scenarios, but the *psycho-physical* laws should not. On the resulting picture, some nomically impossible worlds are causally relevant, i.e. the psychophysically impossible worlds, whereas others are not, i.e., the physically impossible worlds. This measure in effect deals with asteroids suddenly accelerating across the speed of light threshold, whilst still excluding dualist mental causes, thus providing a physicalist-friendly alternative to INTERVENTIONISM IV.

While this could certainly be done *in principle*, I worry that this strategy would fail to establish that anything is *wrong* with INTERVENTIONISM IV. The mental causation objection to dualism is supposed to be that *no* plausible account of causation allows for mental causation in a standard dualist ontology. If we want to retain the high ground in the mental causation debate, we need at least some reason to maintain that there is something wrong with treating

¹⁶ Remember that we disallowed gerrymandering base variables in function of the effect (fn. 13). This assumption plays a central role in physicalist accounts of mental causation as well (e.g., Zhong, 2020a).

the psychophysical laws on a par with the laws of physics and the normality constraints imposed by thermodynamics.

True, INTERVENTIONISM IV imposes strictly weaker requirements than our imagined physicalist friendly alternative. We might insist that such a weakening requires further motivation. Note however, that dualist can help herself to the same strategy that we have been relying when securing higher-level causation. When we exempted physical metaphysical bases, we reasoned that it is no idiosyncrasy of any manipulation of higher-level phenomena that they affect physical metaphysical bases. Similarly, it is no idiosyncrasy of any manipulation of a dualist mental phenomenon that it affects the physical nomic base. After all, it is a matter of nomic necessity that manipulations of dualist mental phenomena come with changes in the physical nomic base. When we excluded thermodynamically abnormal and physically impossible scenarios, we reasoned that there is no real risk in excluding these, as we can in fact always count on finding ourselves in thermodynamically normal and physically possible scenarios. Similarly, we can always count on finding ourselves in psychophysically possible scenarios. After all, these laws are exceptionless and fundamental according to STANDARD DUALISM. The guiding strategy in making adjustments to interventionism has been to isolate those patterns of dependence that are in principle available for reliable manipulation and prediction. Even if STANDARD DUALISM is true, mental phenomena may very well exhibit such patterns of dependence with behavioural effects.

The availability of this strategy to the dualist highlights again how lightweight interventionist accounts of causation are. These accounts do not require that causes produce or necessitate their effects. Instead it suffices that causes and effects exhibit patterns of dependence that can be reliably exploited for the manipulation or prediction of the effect. Moreover, interventionists explicitly argue against the temptation to equate causation with some productive force that underlies these dependences and thus embrace the lightweight nature of

their account (Woodward, 2021, p. 6–8).¹⁷ If we take causation in terms of dependence to be enough for mental causation in physicalist ontologies, we cannot insist that causation in terms of dependence is not enough for mental causation in dualist ontologies (without being very very petty).¹⁸

In short, the problem is not that INTERVENTIONISM IV cannot be denied without delivering faulty results for the physicalist. Instead, the problem is that the reasoning that would take the dualist all the way to INTERVENTIONISM IV runs parallel to the reasoning that we have been relying on to secure higher-level causation and mental causation in physicalist ontologies. On pain of being unprincipled, we cannot simply object to the final step required to attain INTERVENTIONISM IV.

Given this set-up, I see two ways forward. One is to argue that the final step towards INTERVENTIONISM IV makes for problems that are not faced by interventionist accounts that do not allow for dualist mental causation. The other is to reject the interventionist picture altogether. I discuss two variations on the first approach before addressing the second option.

4.2 Too many causes

It is natural to suspect that any account of causation that allows for dualist mental causation must be too permissive. In the case of INTERVENTIONISM IV it is natural to suspect that further relaxing the ‘holding fixed’-condition will cause trouble. After all, that condition was supposed to distinguish causation from spurious correlations due to a common cause. Relaxing it further is bound to blur the distinction (cf. Lewis, 1966).

¹⁷This lightweight approach to causation does not mean that the interventionist gives up on giving an account of what causation *is*. In the cited passage, Woodward separates the ‘minimal’ metaphysical question of what causation is from the ‘ambitious’ metaphysical question of what drives the patterns of dependence and argues that interventionism aims to answer the first question, whilst remaining silent on the second.

¹⁸And if we do not take causation in terms of dependence to suffice, see §4.4.

The crucial adjustment concerns cases involving synchronic nomic necessitation. So if it is to cause troubles, we should expect these to occur in other cases involving synchronic nomic necessitation. Suppose for example that a light bulb emits light because of the electric current in the copper wire connecting the light switch to the bulb. In order to make the causal relation between the current and the emitted light more salient, we can imagine that the intensity of the emitted light is a measure of the electric current. To keep things simple, we can pretend that classical physics is correct. According to Ampère’s law, the electric current flowing through the copper wire in this situation physically necessitates a magnetic field in the surface surrounding the copper wire, and *vice versa*. This appears to be a common cause scenario where one of the causal relations is a synchronic relation of nomic strength: the electric current causes the magnetic field and the emitted light, but the magnetic field does not cause the emitted light.

The objection to aim for is that INTERVENTIONISM IV will count both the magnetic field and electric current as causes of the emitted light. This seems promising. The case seemingly corresponds with the familiar schema in Figure 4. Moreover, the nomic necessitation relation would ascertain that there can be

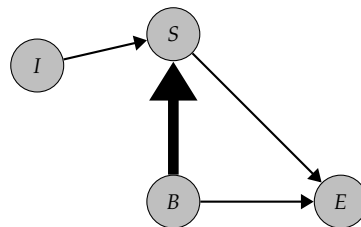


Figure 4: Phenomenon (S) is nomically necessitated by its nomic base (B), both cause target effect (E)

no changes in either the magnetic field or the electric current whilst the other is held fixed. Consequently, the relevant correlations would be trivially robust. On the face of it, the crucial exemption in INTERVENTIONISM IV forces dualists to

conclude that the magnetic field caused the intensity of the emitted light.¹⁹

However, a closer examination of Ampère's law shows that this is not the case. While Ampère's law states that there is a physical equivalence between two variables *in a given situation*, it does *not* state that these two variables are physically equivalent *simpliciter*. That is to say, these equivalences only hold across scenarios where some further factors (other than the fundamental laws of physics) are held fixed.

Here is a more detailed formulation of Ampère's law:

$$\oint_C \vec{B} \cdot d\vec{\ell} = \mu_0 \mu_r I$$

The left hand side refers to the integration of the magnetic field (\vec{B}) along a closed curve (C). The right hand side refers to the product of the permeability of a vacuum (μ_0), the permeability of the surface area (μ_r), and the electric current (I). According to Ampère's law, the magnetic field of the surface is nomically necessitated by these three factors, rather than by the electric current of the copper wire *alone*. There can be changes in the magnetic field in virtue of changes in the medium surrounding the wire rather than changes to the electric current inside the wire. For example, if the wire were surrounded by water or enclosed in rubber, the same current would have given rise to a different magnetic field. The electric current plays a non-redundant part in the nomic base of the magnetic field, but it is, so to speak, not the *full* nomic base of the magnetic field.

The upshot is that the correlation between the magnetic field and the emitted light is, contrary to appearances, not robust. Once the magnetic field is held fixed at any value, nomically possible changes in its full physical base correlate with a change in the emitted light. For example, a change in the permeability of

¹⁹In conversation, some expressed that they find this consequence acceptable. I am not ready to let the dualist off the hook *that* easily.

the medium will nomically require a change in the electric current, which will correlate with a change in the emitted light. Similarly, a change in electric current *is* nomically possible if we keep the electromagnetic force constant. It will just require a change in the permeability of the medium as well. Such changes in the permeability and the current *will* correlate with changes in the emitted light. Note that none of these changes require nomically impossible scenarios or thermodynamically abnormal realizations. Consequently, the correlation between the magnetic field and the emitted light is *not* sufficiently robust and INTERVENTIONISM IV gets it just right: the electric current causes the emitted light, and the magnetic field does not.

We can hold out hope that this is a peculiarity of the example, but I am afraid the problem runs deeper. Consider how our intuitions change if we assume that the correlation between changes in the magnetic field and the emitted light *does* appear robust. Rather than deciding that INTERVENTIONISM IV is wrong, I'd be inclined to conclude that the magnetic field *is* the cause of the emitted light. So even in other cases involving synchronic nomic necessitation, INTERVENTIONISM IV appears to provide a principled distinction between causation and confounded correlations.

4.3 Exclusion arguments

Many argue that, if PHYSICAL COMPLETENESS is true, then non-physical phenomena can only cause physical effects by way of 'overdetermination'. If this reasoning is correct, and STANDARD DUALISM is true, then mental phenomena can only be overdetermining causes. The upshot would be that, even if INTERVENTIONISM IV delivers dualist mental causation, it can only do deliver it in the form of overdetermination. A consequence that is widely taken to be unacceptable.

Here is the reasoning a bit more precisely. The driving idea is:

Exclusion For any two metaphysically distinct phenomena C_1 and

C_2 , and any effect E : if both C_1 and C_2 occur at t and C_1 is physically sufficient for E 's occurrence at $t + x$, then C_2 cannot be a cause of E , unless it is a case of overdetermination.²⁰

Together with the PHYSICAL COMPLETENESS component of STANDARD DUALISM, EXCLUSION excludes dualist mental phenomena from being overdetermining causes. Dualist pain is metaphysically distinct from the co-occurrent physical phenomenon that physically necessitates the subsequent wincing. So, either the pain is not a cause of the wincing, or it is a case of overdetermination. So states EXCLUSION.

Making all cases of mental causation also cases of overdetermination is widely taken to be problematic. Overdetermination cases require an *ad hoc* back-up construction to afford causal status to two independent causal mechanisms that do not require one another's workings to bring about the effect. For example, if a victim is killed by two simultaneous gunshots by different gunmen, each of which would have brought about the victim's death in absence of the other, then the victim's death was overdetermined. While such set-ups are certainly possible, it would be all too baroque to posit them as a widespread feature of reality.²¹ It would be bad news for dualists if mental causation requires such widespread overdetermination.

While this is a popular line of argument, it is hard to see how it can pose a real challenge for the current account. According to INTERVENTIONISM IV, dualist mental causes require no baroque back-up mechanisms. Their causal status just derives from the correlation patterns being as they are. Moreover, mental causes would fail the standard test for overdetermining causes devised by Bennett (2003; 2008). According to this test, my pain (Pa) and its underlying physical phenomenon (Ph) overdetermine my wincing Wi if both the closest Ph

²⁰This formulation does not call C_1 a cause of B to avoid worries about causal sufficiency. I motivate this manoeuvre elsewhere (Vaassen, 2021a).

²¹Cf., Papineau (2002, §5) and Bennett (2003, p. 475).

and not-*Pa* world and the closest *Pa* and not-*Ph* world still contain *Wi*. This is not the case on the dualist picture. In '*Pa* and not-*Ph*'-worlds, *Ph* is excised cleanly from reality without replacing it with a similar phenomenon. In such worlds, I would not wince. Hence, it is not a case of overdetermination.

Bennett goes on to argue that the dualist requires a credible account of causation that allows *Pa* to still be a cause of my wincing, despite the "if *Pa* and not-*Ph*, then *Wi*" counterfactual being false. INTERVENTIONISM IV does just that.²² According to this account, the '*Pa* and not-*Ph*'-world plays no role in determining *Pa*'s causal profile. Instead of looking at that counterfactual world, we should look first at the nomically possible world in which *Pa* is absent (and thus *Ph* is absent as well), and then look at those worlds where *Pa* has different underlying physical phenomena to ascertain that the *Pa*-*Wi* correlation is robust. *Pa* can pass both these tests without passing the "if *Pa* and not-*Ph*, then *Wi*" test.

If INTERVENTIONISM IV is true, then EXCLUSION is false. And so the reasons motivating INTERVENTIONISM IV also motivate a denial of EXCLUSION. If we want to hold onto the idea that STANDARD DUALISM does not allow for mental causation, we need to find fault with INTERVENTIONISM IV. Our first two attempts at doing so were unsuccessful. Let us try one more.

4.4 Oomph!?

One could object to INTERVENTIONISM IV by objecting to its interventionist origins. Interventionist accounts belong to the broader class of *dependence* views, which define causation in terms of patterns of counterfactual dependence. Such accounts are often opposed for mistaking the symptoms of a causal relation, i.e., counterfactual dependence, for the real thing, i.e., the process of production or necessitation that gives rise to the counterfactual dependence. Dependence views have, so to speak, taken the *oomph* out of causation. If this is a mistake,

²²It can thus be seen as a friendly elaboration on List and Stoljar's (2017) response to exclusion arguments, in light of my own concerns with their response Vaassen (2021b).

then INTERVENTIONISM IV is built on shaky ground.

The objection from *oomph* is worth taking seriously. It is intuitively compelling and has strong roots in debates about both causation in general and mental causation in particular (e.g. Kim, 2007; Ney, 2009; Dowe, 2009). The problem is that it is unlikely to do us physicalists any good. Theories of causation that do conserve the *oomph* in causation are so demanding as to exclude mental causation within standard physicalist frameworks as well. Let us look at two examples.²³

According to production views on causation, causes *produce* their effects by transferring a conserved quantity on their effects (e.g. Dowe, 2000). One direct consequence of such views is that absences cannot be causes, as they fail to transfer a conserved quantity on their effects (cf., Schaffer, 2000; Dowe, 2001). The problem is that absences play a crucial role in connection between phenomena in the brain and physical behavior. In particular, the mechanism underlying muscle contraction relies on the causal role of tropomyosin absences to connect signals from the brain to bodily movements (Schaffer, 2004, p. 200). I refer the reader to Russo (2016) for a more detailed discussion, but the upshot is simple: there is no suitably productive process connecting our brain phenomena to our bodily behaviour. If the production view is correct, this means that there is no causal connection between brain phenomena and bodily behaviour either. On the plausible assumption that brain phenomena would be the relevant realizers or reduction bases for mental phenomena in physicalist ontologies, it follows that there is no mental causation in physicalist ontologies.²⁴

Similarly, theories that require causes to *necessitate* their effects stand little chance of securing mental causation in physicalist theories. Necessitating an effect, it turns out, is a task that only the most fine-grained physical phenomena can achieve (cf., Loewer, 2007b). As discussed in §3.1, macrophysical phenom-

²³See also Woodward (2008) for a similar line of reasoning.

²⁴*Mutatis mutandis*, such concerns transfer to other more oomphy views that do not allow for causation by absence, such as the powers view employed in Gibb (2013).

ena can be realized in thermodynamically abnormal ways, such that their future is highly unpredictable. For example, my C-fibers firing could be realized in such a way as to change me into a mule, rather than making me wince. The fundamental laws of nature only relate phenomena individuated at the finest grain, such as leptons and quarks having spin or mass, rather than coarse-grained phenomena such as C-fibers firing and other neural activity. Consequently, the neural phenomena that underly our mental phenomena are unlikely candidates for being causes on a necessitation view. Again, even the sensible physicalist view is left without mental causation

The standard way of assuaging the demanding nature of oomphy views is to distinguish between real, full-blooded causation and a less demanding notion of causal explanation or causal relevance. The general idea is that phenomena that are standardly considered causes but do not meet the oomphy criteria can still feature in respectable causal explanations.(e.g. Jackson and Pettit, 1990; Dowe, 2001; Gibb, 2013) So, even if our mental phenomena do not cause our behaviour, they could still causally explain our behaviour. The problem is that this approach would still put the physicalist in the same boat as the dualist. Dualists can just rehearse the arguments provided above and present them as arguments about causal explanation rather than genuine causation. The upshot would be that both physicalists and dualists cannot have genuine, full-blooded mental causation, but they can allow for mental phenomena to causally explain our behaviour.

In short, oomphy views are a tide that sinks all boats. Both dualists and physicalists would lose out on mental causation and both could settle for causal explanation involving mental phenomena instead. So, even if we decide to adopt an oomphy account, mental causation worries do not provide the physicalists with the dialectical advantage we thought it did.

5 Conclusion

The standard objection to STANDARD DUALISM is that it cannot allow for mental causation. I have argued that there is a credible account of causation that allows for dualist mental causation. Broadly speaking, the account is motivated by the fact that higher-level phenomena are taken to be causes in virtue of exhibiting sufficiently robust patterns of correlation, and dualist mental phenomena can feature in such patterns as well. The upshot is that physicalists are in need of better objections to STANDARD DUALISM. Believe me. I'm not particularly happy about it either.

Acknowledgements

I am heavily indebted to Pär Sundström, Torfinn Huvenes, Gunnar Björnsson, Andreas Stokke, Thomas Kroedel, Mathias Frisch, and David Papineau for help with developing the overall argument. I am grateful to Ethan Nowak, Karen Bennett, Ezra Rubenstein, Andrew Rubner, and Audrey Powers for comments on later versions of the text. Insightful comments by two anonymous referees and the editor at this journal have helped me situate the main claims of the paper in broader debates about the nature of causation and the importance of mental causation. This research was funded by a Swedish Research Council International Postdoc Grant [grant number 2019-0638].

References

- Albert, D. Z. (2015). *After Physics*. Harvard University Press.
- Baumgartner, M. (2010). Interventionism and epiphenomenalism. *Canadian Journal of Philosophy*, 40(3):359–383.

- Bealer, G. (2007). Mental causation. *Philosophical Perspectives*, 21(1):23–54.
- Bennett, K. (2003). Why the exclusion problem seems intractable, and how, just maybe, to tract it. *Noûs*, 37(3):471–497.
- Bennett, K. (2008). Exclusion again. In Kallestrup, J. and Hohwy, J., editors, *Being Reduced*, pages 280–307. Oxford University Press.
- Blanchard, T. (2020). Explanatory abstraction and the goldilocks problem: Interventionism gets things just right. *British Journal for the Philosophy of Science*, 71(2):633–663.
- Campbell, J. (2010). Control variables and mental causation. *Proceedings of the Aristotelean Society*, 110(October 2009):15–30.
- Chalmers, D. J. (1996). *The Conscious Mind*. Oxford University Press.
- Chalmers, D. J. (2013). Panpsychism and panprotopsychism. *Amherst Lecture in Philosophy*, 8.
- Dorr, C. (2016). Against counterfactual miracles. *Philosophical Review*, 125(2):241–286.
- Dorst, C. (2022). Why do the laws support counterfactuals? *Erkenntnis*, 87:545–566.
- Dowe, P. (2000). *Physical Causation*. Cambridge University Press.
- Dowe, P. (2001). A counterfactual theory of prevention and ‘causation’ by omission. *Australasian Journal of Philosophy*, 79(2):216–226.
- Dowe, P. (2009). Causal process theories. In Beebe, H., Hitchcock, C., and Menzies, P., editors, *The Oxford Handbook of Causation*. Oxford University Press.
- Franklin-Hall, L. R. (2016). High-level explanation and the interventionist’s ‘variables problem’. *British Journal for the Philosophy of Science*, 67(2):553–577.

- Gallow, J. D. (2015). The emergence of causation. *Journal of Philosophy*, 112(6):281–308.
- Gibb, S. C. (2013). Mental causation and double prevention. In Gibb, S. C., Lowe, E. J., and Ingthorsson, R. D., editors, *Mental Causation and Ontology*, pages 193–213. Oxford University Press.
- Gibb, S. C. (2015). VIII—Defending dualism. *Proceedings of the Aristotelian Society*, 115(2pt2):131–146.
- Goff, P. (2017). *Consciousness and Fundamental Reality*. Oxford University Press.
- Goodman, J. (2015). Knowledge, counterfactuals, and determinism. *Philosophical Studies*, 172(9):2275–2278.
- Ismael, J. T. (2016). *How Physics Makes Us Free*. Oxford University Press.
- Jackson, F. and Pettit, P. (1990). Causation and the philosophy of mind. *Philosophy and Phenomenological Research*, 50(n/a):195–214.
- Kaiserman, A. (2020). Interventionism and mental surgery. *Erkenntnis*, 85(4):919–935.
- Khawaja, J. (2022). Interactionist zombies. *Synthese*, 200.
- Kim, J. (2005). *Physicalism, or something near enough*. Princeton University Press.
- Kim, J. (2007). Causation and mental causation. In McLaughlin, B. P. and Cohen, J. D., editors, *Contemporary Debates in Philosophy of Mind*, pages 227–242. Blackwell.
- Kroedel, T. (2015). Dualist mental causation and the exclusion problem. *Noûs*, 49(2):357–375.
- Kroedel, T. (2020). *Mental Causation: A Counterfactual Theory*. Cambridge University Press.

- Lewis, D. (1966). An argument for the identity theory. *Journal of Philosophy*, 63(1):17–25.
- List, C. and Menzies, P. (2009). Non-reductive physicalism and the limits of the exclusion principle. *The Journal of Philosophy*, 106(9):475–502.
- List, C. and Stoljar, D. (2017). Does the exclusion argument put any pressure on dualism? *Australasian Journal of Philosophy*, 95(1):96–108.
- Loewer, B. M. (2007a). Counterfactuals and the second law. In Price, H. and Corry, R., editors, *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, pages 293–326. Oxford University Press.
- Loewer, B. M. (2007b). Mental causation, or something near enough. In McLaughlin, B. P. and Cohen, J. D., editors, *Contemporary Debates in Philosophy of Mind*, pages 243–64. Blackwell.
- Lowe, E. J. (1996). *Subjects of Experience*. Cambridge University Press.
- Lowe, E. J. (1999). Self, agency, and mental causation. *Journal of Consciousness Studies*, 6(8):225–239.
- Lowe, E. J. (2008). *Personal Agency: The Metaphysics of Mind and Action*. Oxford University Press.
- Lycan, W. G. (2009). Giving dualism its due. *Australasian Journal of Philosophy*, 87(4):551–563.
- McGrath, S. (2005). Causation by omission: A dilemma. *Philosophical Studies*, 123(1-2):125–148.
- Ney, A. (2009). Physical causation and difference-making. *British Journal for the Philosophy of Science*, 60(4):737–764.
- Papineau, D. (2002). *Thinking about Consciousness*. Oxford University Press.

- Papineau, D. (2013). Causation is macroscopic but not irreducible. In Gibb, S., Lowe, E. J., and Ingthorsson, R., editors, *Mental Causation and Ontology*, pages 126–151. Oxford University Press.
- Papineau, D. (2022). The statistical nature of causation. *The Monist*, 105(2):247–275.
- Raatikainen, P. (2010). Causation, exclusion, and the special sciences. *Erkenntnis*, 73(3):349–363.
- Russo, A. (2016). Kim’s dilemma: Why mental causation is not productive. *Synthese*, 193(7):2185–2203.
- Schaffer, J. (2000). Causation by disconnection. *Philosophy of Science*, 67(2):285–300.
- Schaffer, J. (2004). Causes need not be physically connected to their effects: The case for negative causation. In Hitchcock, C. R., editor, *Contemporary Debates in Philosophy of Science*, pages 197–216. Basil Blackwell.
- Touborg, C. T. (2022). Relativizing proportionality to a domain of events. *Synthese*, 200(2):1–20.
- Vaassen, B. (2019). *Causal After All : A Model of Mental Causation for Dualists*. PhD thesis.
- Vaassen, B. (2021a). Causal exclusion without causal sufficiency. *Synthese*, 198:10341–10353.
- Vaassen, B. (2021b). Dualism and exclusion. *Erkenntnis*, 86(3):543–552.
- Vaassen, B. (2022). Halfway proportionality. *Philosophical Studies*, (9):1–21.
- Walter, S. (2010). Taking realization seriously: No cure for epiphobia. *Philosophical Studies*, 151(2):207–226.

- Won, C. (2021). Mental causation as joint causation. *Synthese*, 198(5):4917–4937.
- Woodward, J. (2003). *Making Things Happen: a Theory of Causal Explanation*. Oxford University Press.
- Woodward, J. (2008). Mental causation and neural mechanisms. In Kallestrup, J. and Hohwy, J., editors, *Being Reduced*, pages 218–263. Oxford University Press.
- Woodward, J. (2014). A functional account of causation; or, A defense of the legitimacy of causal thinking by reference to the only standard that matters—usefulness. *Philosophy of Science*, 81(5):691–713.
- Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, 91(2):303–347.
- Woodward, J. (2018). Explanatory autonomy: The role of proportionality, stability, and conditional irrelevance. *Synthese*, pages 1–29.
- Woodward, J. (2020). Causal complexity, conditional independence, and downward causation. *Philosophy of Science*, 87(5):857–867.
- Woodward, J. (2021). *Causation with a human face: Normative theory and descriptive psychology*. Oxford University Press.
- Yablo, S. (1992). Mental causation. *Philosophical Review*, 101(2):245–280.
- Zhong, L. (2014). Sophisticated exclusion and sophisticated causation. *Journal of Philosophy*, 111(7):341–360.
- Zhong, L. (2020a). Intervention, fixation, and supervenient causation. *Journal of Philosophy*, 117(6):293–314.
- Zhong, L. (2020b). Taking emergentism seriously. *Australasian Journal of Philosophy*, 98(1):31–46.