Synthese (2023) 202:53
https://doi.org/10.1007/s11229-023-04282-4

ORIGINAL RESEARCH

Emergent mental properties are not just double-preventers

Andrei A. Buckareff[1] · **Jessica Hawkins**[1]

Received: 1 December 2022 / Accepted: 20 July 2023 / Published online: 8 August 2023

Abstract
We examine Sophie Gibb's emergent property-dualist theory of mental causation as double-prevention. Her account builds on a commitment to a version of causal realism based on a powers metaphysic. We consider three objections to her account. We show, by drawing out the implications of the ontological commitments of Gibb's theory of mental causation, that the first two objections fail. But, we argue, owing to worries about cases where there is no double-preventive role to be played by mental properties, her account, which solely affords mental properties a double-preventive role, is incomplete and vulnerable to a causal exclusion objection. We propose a friendly modification to her theory of mental causation that is consistent with her theory's ontological commitments. Specifically, we sketch an account on which mental properties have a more pronounced causal-structuring role that is not exhausted by the role Gibb assigns them as double-preventers. The result is a novel emergentist theory of mental causation.

Keywords Mental causation · Causation · Mind · Powers · Properties · Dualism · Emergentism · Metaphysics

1 Introduction

The problem of mental causation has been regarded as especially problematic for dualist proposals in the metaphysics of mind. In the recent literature, novel proposals have been offered by some proponents of dualism in response to challenges from critics.[1] Sophie Gibb has offered an original account of the causal relevance of irreducible emergent mental properties in a series of papers (Gibb, 2013, 2015a, 2015b, 2015c, 2019). Specifically, Gibb has argued that mental properties are causally relevant by

[1] See, especially, the work of E. J. Lowe (2006, 2008).

✉ Andrei A. Buckareff
andrei.buckareff@marist.edu

Jessica Hawkins
jessica.hawkins1@marist.edu

[1] Department of Philosophy and Religious Studies, Marist College, Poughkeepsie, NY 12601, USA

being double-preventers. "Double prevention occurs when an event that would prevent another event from having a certain effect is itself prevented from doing so" (Gibb, 2013, p. 198). Gibb's proposal allows the property-dualist a way around the causal exclusion problem. In particular, if Gibb's theory of mental causation as double-prevention works, then the emergent property-dualist can accept all of the assumptions forwarded by their reductionist interlocutors in versions of the causal exclusion problem without contradiction.

Our primary goal in this article is to consider whether Gibb's theory of mental causation can deliver the truthmakers for representing ourselves as agents who produce outcomes in the world. We do this by considering some actual and potential objections to her theory, filling in some gaps in Gibb's account that have left it vulnerable to criticism. The end result is a theory of mental causation for emergent property-dualists that affords a more expansive role for mental properties beyond what Gibb assigns them as double-preventers. Importantly, the account we sketch does not involve the addition of any ontological commitments beyond those implied by Gibb's theory of mental causation as articulated in her work.

We will proceed as follows. We will first discuss Gibb's formulation of the causal exclusion problem, followed by a summary of her theory of mental causation as double-prevention. We will then consider and respond to a worry raised by Kim Davies (2016) related to whether Gibb's theory of mental causation is committed to a kind of pre-established harmony at the mental level. This is followed by a summary and response to some worries coming from the work of Jaegwon Kim and Stephen Mumford and Rani Anjum. We finally consider a more challenging worry coming from Davies involving cases where there is no double-prevention. This last challenge we take to be the most formidable, pointing to some gaps in Gibb's theory that we attempt to fill. The end result is a more comprehensive theory of mental causation for the emergent property dualist. The account expands on Gibb's account, drawing out the implications of some of the ontological commitments of her theory that she leaves underdeveloped or unarticulated. But in doing so, we argue for a more pronounced role for emergent mental properties beyond the role Gibb assigns them as double-preventers.[2]

2 The causal exclusion problem

The causal exclusion problem in the metaphysics of mind can be traced back to the problem of explanatory exclusion articulated in Norman Malcolm's paper, "Conceivability and Mechanism" (1968). Malcolm was chiefly concerned with the relationship between purposive psychological explanations and non-purposive mechanistic neurophysiological explanations. Jaegwon Kim (1989) refined the general line of reasoning, offering the blueprint for subsequent statements of exclusion arguments. Kim shifted

[2] While both of the authors of this article regard Gibb's strategy and her general theory of mental causation as attractive, providing a promising framework for developing a solution to the causal exclusion problem for emergent property dualists, it is not the case that both authors accept dualism of any sort. One of the authors (Andrei Buckareff) favors a version of neutral monism that is best described as a version of panprotopsychism on which all properties are proto-phenomenal powers. That said, both authors regard the account sketched here to be the best way to fix Gibb's account and think it is the best option for those who endorse theories that are committed to mental properties being strongly emergent.

the focus away from being chiefly epistemological, focusing on explanations, to being primarily about metaphysics, specifically over what causal role is left for mental properties if they are ontologically dependent upon, while being ontologically irreducible to, physical properties (Kim, 2005).

Gibb's formulation of the problem mostly follows the template given by Kim. She presents the causal exclusion problem as resulting from a putatively inconsistent set of "four claims [that seem] individually plausible" (Gibb, 2013, p. 192).

(1) Relevance: Mental events are causally relevant in the physical domain.
(2) Closure: Every physical event contains only other physical events in its transitive causal closure.
(3) Exclusion: As a general rule, events are not causally overdetermined.
(4) Distinctness: Mental events are not physical events. (Ibid., 193)

In the remainder of this section, we will first discuss each of the elements of the set and then take up why they are regarded as being inconsistent.

First, consider (1) *Relevance.* Assuming causal realism, following Gibb, we assume that something p is causally relevant to the production of an outcome q if q counterfactually depends upon p (2013, p. 202). Something can be causally relevant in *at least* two different ways.[3] It may be causally relevant by causally *producing* an outcome or causally *enabling* an outcome. Causal production is a relatively straightforward notion. Causal enabling is less so. For our purposes, we take something p to causally enable an outcome q if and only if (a) p is a non-productive element in a causal process that generates q and (b) q would not have occurred absent the activity of p.

Regarding (2) *Closure,* Gibb's definition of *Closure* echoes E. J. Lowe's formulation of the closure principle which invokes the notion of a physical event having "only other physical events in its *transitive causal closure*" (Gibb, 2013, p. 193, emphasis added). Lowe articulates what is meant by 'transitive causal closure' as follows.

> By the 'transitive causal closure' of an event P, I mean the set of events consisting of the immediate causes of P, the immediate causes of those causes, the immediate causes of *those* causes … and so on: in short, the set which includes every event which stands in the ancestral of the 'immediate cause' relation to P. The implication . . . , then, is that the immediate causes of all physical events are always and only other physical events. (Lowe, 2000, pp. 581–582)

Thus, assuming transitive causal closure, an agent's arm rising would have a physical cause as its immediate cause.

Unfortunately, Gibb does not offer a very informative explication of (3) *Exclusion.* We will simply assume Kim's formulation.

> *Exclusion.* No single event can have more than one sufficient cause occurring at any given time—unless it's a genuine case of causal overdetermination. (Kim, 2005, p. 42).

As we shall see, the problem of mental causation for the property dualist rests largely on the assumption of the foregoing *Exclusion* and *Closure* principles.

[3] We specify some further ways a property can be causally relevant in Sect. 6, below.

As for (4) *Distinctness,* this is the core assumption of all dualists, including property dualists. Mental events are not physical events. Going further, the constitutive properties of mental events are not physical properties.

We can now state the problem as Gibb presents it. If mental events are not physical events and are causally relevant in the physical domain, there are worries about *Closure* being violated (so (1) and (4) are in tension with (2)). If we accept (1), (2), and (4), then we cannot also accept that there is no overdetermination (3). Gibb elaborates on the problem as follows.

> The apparent inconsistency of these claims gives rise to the problem of mental causation. Responses to this problem typically provide reasons to reject one of the claims. Indeed, the various positions in the mental causation debate can, to a large extent, be distinguished by the claim that they reject. Hence, eliminativism and epiphenomenalism both reject *Relevance*. Interactive substance dualism and anti-physicalist forms of property dualism typically deny *Closure*. Most forms of non-reductive physicalism deny (or disambiguate) *Exclusion*. Finally, psychophysical reductionism rejects *Distinctness*. (Gibb, 2013, p. 193)

Gibb's strategy, as we shall make clear below, is a version of compatibilism about mental causation. She aims to provide a dualist theory of mental causation that is consistent with accepting (1)–(4). If successful, this is a promising strategy for those who reject the reducibility of mental properties to physical properties. We will next outline the central features of Gibb's theory of mental causation.

3 Gibb's theory of mental causation

In presenting her theory of mental causation as double-prevention, Gibb has adopted a neutral stance in some of her publications between substance-dualism and property-dualism (2015b, pp. 626–627, n. 1). Of course, the former entails the latter and, as Susan Schneider (2012) has forcefully argued, the latter may commit one to the former view.[4] That being said, elsewhere, Gibb has expressed a commitment to a view of mental *entities* as ontologically dependent upon physical *entities*. Given the context, with her emphasis upon mental properties and events, we assume she means to pick out events and their constitutive properties by the term 'entities' (Gibb, 2019).

Gibb takes mental properties to be strongly emergent from physical properties. Hence, mental properties are fundamentally novel and irreducible to physical properties. On the account of emergence Gibb assumes, the ontological dependence in question is a form of *causal dependence* (Gibb, 2015a, p. 145, n. 7; Gibb, 2015c, p. 85). An emergent mental event m_1 and a concurrent neural event n_1 have a common cause in a neural event n_0.

[4] Specifically, Schneider (2012) has argued that property-dualism commits one to either Cartesian substance dualism or a dualism of physical and hybrid substances. The argument is based on the assumption that properties are ways that substances are. Thus, if a substance has both mental and physical properties, one of the ways it *is* is characterized by the mental properties. Hence, we have a hybrid substance that is not purely physical. Schneider suggests that the property dualist endorse the hybrid view. We expect that Gibb would be sympathetic to this claim given that she endorses the view that properties are borne by and are ways that substances are (see Gibb 2017).

Gibb endorses a version of powers realism about properties. She uses 'disposition' and 'power' interchangeably (Gibb, 2015a, p. 140; Gibb, 2015b, p. 632; Gibb, 2015c, p. 77).[5] Going further, she assumes a version of pan-dispositionalism on which "all intrinsic properties are dispositional, where a property is dispositional if, solely in virtue of instantiating it, a substance possesses a certain power" (Gibb, 2013, p. 196). She remains neutral between the various forms of powers realism. In particular, she adopts a neutral stance in the "pure powers" versus "powerful qualities" debate (Gibb, 2015c, p. 77, n. 8).[6] Gibb, otherwise, accepts the standard litany of ontological commitments of powers realism, including the following. She denies that dispositional properties are grounded in categorical properties and rejects a conditional analysis of 'disposition'. Finally, she denies that the manifestation of a disposition is the same thing as its possession. An object can possess a disposition that it never manifests.

Gibb's assumptions about the metaphysics of causation are not articulated in much detail. Given that she takes her account to simply be the one articulated by Martin (2008),[7] it may not be surprising that she does not go into much detail about what is going on in causal processes. Gibb takes powers to be directed at mutually manifesting with various reciprocal manifestation partners. Hence, when objects whose powers are directed at mutually manifesting interact, the powers in question are manifested. The mutual manifestation of reciprocal manifestation partners, together, causally generate an outcome. Therefore, all of the manifesting powers contribute to an outcome (Gibb, 2013, pp. 197–198). Effects are polygenic (Molnar, 2003).

The literature by proponents of powers realism on the topic of double-prevention is a small subset of work on the powers theory of causation (see, e.g., Anjum and Mumford, 2009; Mumford & Anjum, 2011). "In cases of *double* prevention a disposition that is disposed to prevent the manifestation of another disposition, is itself prevented from doing so by the presence of a third disposition" (Gibb, 2013, p. 200). Consider the following example. Someone ingests a poison that will take ten minutes to kill them and is given an antidote that would mask the power of the poison to kill them. They subsequently drink some water that was laced with a compound that will neutralize the antidote, allowing the poison to kill them. The compound's power is a double-preventer. Framed in terms of dispositions and their manifestations, a disposition that is a double-preventer, when manifested, will block the manifestation of a particular disposition of an object that would block the manifestation of another disposition. In the case of the poisoning, the manifestation of the compound's power to neutralize blocks the manifestation of the antidote whose manifestation would have blocked the manifestation of the poison's disposition to kill.

The double-preventer is not an efficient cause. That is, the causal relevance of a double-preventer in a causal process is not owed to the double-preventer directly contributing to the causal production of an outcome. Rather, the double-preventer permits or enables "a cause to bring about an effect." Gibb maintains that it "is no less

[5] See Buckareff 2022a for a powers realist argument against identifying powers with dispositions and against taking 'power' and 'disposition' to be interchangeable (powers provide the truthmakers for disposition-ascriptions).

[6] See Taylor 2018 for an argument to the effect that there is no real difference between a pure powers view and at least certain accounts of powers as powerful qualities.

[7] Martin's discussion, while quite rich, is very short (spanning pages 48–51 of his 2008).

important than the role of the cause" (Gibb, 2013, p. 202). Elsewhere, she refers to a double-preventer as an "*enabling event*" (Gibb, 2015b, p. 636).

Gibb takes enabling events to be causally relevant to the outcomes they permit in virtue of the outcomes being counterfactually dependent upon the enabler. She writes:

> Now, although enabling events are not causes of the events that they enable to be caused, I would suggest that the former events are causally relevant to the latter events, and not just in a merely explanatory sense. In causal situations in which enabling events are involved, for the effect to be brought about, in addition to the complete cause, a further event must occur whose role is to enable the causal relation to take place. Clearly, this role is an objective one, not a merely explanatory one. Furthermore, I can see no good reason for thinking that the role of an enabling event is any less important than the role of a cause in accounting for an effect's existence. (Gibb, 2015b, p. 637)

We will return to how Gibb characterizes enabling events below. In particular, her way of characterizing enabling events strikes us as more modest than it needs to be, especially given her commitment to powers realism and a powers theory of causation.

This brings us to Gibb's theory of mental causation as double-prevention. Mental events function as double-preventers. So they are "causally relevant in the physical domain, not because they cause physical events, but because they enable physical events to be caused" (Gibb, 2015a, p. 142). How does this work? Assume the truth of dualism about mental events and properties. Hence, mental events are not identical with physical events. Now, consider the case of Fred. Fred has been sitting for some time with a cat sleeping on his lap. Fred desires that the cat remain on his lap, but he is also quite uncomfortable and desires that his posture change, which will result in the cat jumping off his lap. Following Gibb, take 'n_1' to pick out the neural event that would serve to trigger neural event 'n_2' that would result in the muscle contractions constitutive of Fred's moving, which we will designate by 'b_1'. We will take 'm_1' to refer to his desire to move and 'm_2' to refer to Fred's desire that the cat remain on his lap. Assuming that m_1 is the stronger desire, Gibb takes it to be the case that m_1 prevents the manifestation of the constitutive power of m_2 (Gibb, 2015a, p. 143). If it were not the case that m_1 were stronger than m_2, then m_2 would be manifested, resulting in Fred remaining still. So, according to Gibb, "m_1 prevents m_2 from preventing n_2 from causing b_1" (Gibb, 2015a, p. 143; see Fig. 1).

m_1 is causally relevant, not in virtue of causally *producing* b_1, but in virtue of causally *enabling* n_2 to cause b_1. Gibb maintains thatn_2 is the complete cause of b_1,

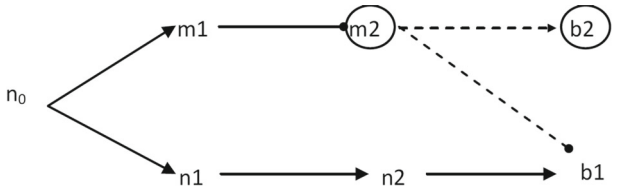

Fig. 1 Mental causation as double-prevention

and n_1 is the complete cause of n_2. m_1 is a double-preventer that, again, causally enables n_2 to bring about b_1. On this account, we have irreducible mental events and their constitutive properties being causally relevant without any violation of *Closure* or *Exclusion*.

> Unlike standard dualist models of psychophysical causal relevance, the double prevention model denies that the causal role of mental events in the physical domain is to cause some neural event or set of neural events which ultimately gives rise to some bodily movement. Rather than causing any neural event, the causal role of mental events in the physical domain is to enable neural events to give rise to bodily movements. (Gibb, 2015a, pp. 143–144).

Gibb's strategy strikes us as promising. Whether or not it is finally tenable depends upon how well it can weather objections from critics. We will argue that Gibb's theory will only succeed if certain modifications and additions are made to the existing account. As we shall see, each objection presents an opportunity to flesh out the implications of the account or else to revisit some of the tools provided by Gibb's assumptions that she leaves underspecified.

4 Preserving closure

Kim Davies (2016) has argued that, on Gibb's account, the *Closure* hypothesis "is maintained only on the assumption of a pre-ordained harmony between preventing and double-preventing mental events which stretches plausibility to its outer limits and for which there is no independent support" (Davies, 2016, p. 42). Davies takes the central problem to be that Gibb's proposal still leaves it as a metaphysical possibility that mental events and their constitutive powers are causally efficacious in the physical realm. "These dispositions are... never manifested as they are always either countermanded by a double preventer desire or are in accord with what would anyway be the physical train of events...." (Davies, 2016, p. 39). Davies argues that Gibb's model assumes a picture on which mental causal interventions into the physical are possible, but never actual. There are real unmanifested causal powers constitutive of desires that, if manifested, could result in causal interventions at the physical level (Davies, 2016, pp. 40–41). Neurophysiological chains give rise to emergent mental events that "ride shotgun as it were, and defend the causal chain against any potential hold-ups perpetrated by would-be intervening maverick desires m_2" (Ibid., p. 40).

Davies' objection appears at times to rest on a misunderstanding of the metaphysical framework within which Gibb is working. First, Davies does not tell us how he is thinking about laws of nature. But in the only place where he discusses laws of nature (Ibid., p. 39–40), he offers remarks that suggest he assumes laws of nature are governing. He writes as follows about the laws associated with the possible interventions that would violate *Closure*:

> These interventions, were they actually to take place, are ***governed*** by irreducibly psychophysical laws. These laws however are never implemented, as the interventions never do take place. (Ibid., p. 40, emphasis added)

In the same essay by Gibb that Davies discusses, Gibb notes that, on "the powers theory of causation, laws are nothing more than generalized claims about causal relations and causal relations are to be accounted for in terms of powers" (Gibb, 2013, p. 206). In other words, the laws of nature *describe* the actual and potential manifestations of the powers of objects in the world. The laws themselves *do not govern*. And while the constitutive properties of some mental events are powers that are directed at preventing physical events, the powers in question are never actually manifested. The upshot is that "the conservation laws are never actually violated" (Ibid., p. 206). But this is not owing to laws external to the powers themselves that govern their manifestations.

Davies may allow that the foregoing is all fine and good. His use of language that implies that laws are governing may be owing to an oversight or a momentary lapse into using language commonly deployed in discussions of laws of nature. Whether laws are governing or not, he may maintain that what Gibb describes is implausible. It is still the case that mental powers *can* be manifested in a way that results in the violation of *Closure* but are never manifested in this way. On Gibb's account, it is metaphysically possible that *Closure* is violated, given that the agent in question has a desire whose constitutive power is directed at a manifestation that would violate causal closure. Conveniently, this violation of *Closure* will never happen. Why think this?

In brief, *Closure* will not be violated because the strongest power always wins. The manifestation of a mental power would never result in the violation of *Closure*. Why this is so will require some explanation and will also require that we go beyond what Gibb says about the metaphysics of causation. The way Gibb represents causal processes is perhaps the source of confusion that leads those like Davies to offer the kind of objection he offers.

Gibb follows the common trend of representing causal processes using neuron diagrams and referring to events as causal relata (we even follow Gibb and use a neuron diagram in Sect. 6 below). Neither is advisable if we assume an ontology of causal powers and a metaphysics of causation that builds upon it.[8] The reason is that representing causal processes in this sort of way fails to deliver an accurate representation of them. (Thus, the limitations of neuron diagrams as a way of modeling causal relations needs to be acknowledged.) That said, nothing we say in this section will be inconsistent with her other commitments. Rather, we will simply be trying to flesh out some of Gibb's commitments in more detail than she does.

Recall that n_1 and m_1 share a common cause, namely, n_0. Given that Gibb explicitly endorses a powers theory of causation, we need to think about what is going on when n_0 causes n_1 and m_1. A widely held metaphysics of events is Jaegwon Kim's (1976) on which an event is the exemplification of a property by an object at a time.[9] Some find this theory of events as inadequate given that events should be understood as being a bit more dynamic. Hence, taking Kim's analysis as a starting point, some follow Lawrence Lombard (1979, 1998) and take events to occur when there is a change in properties being exemplified by an object. Regardless of which theory of events

[8] For more on the activation of causal powers in causal processes in the context of exercising intentional agency, see Buckareff (2018, 2022b) and Aguilar and Buckareff (2022).

[9] Gibb explicitly endorses a Kimean metaphysics of events. She takes both causes and effects to be Kimean events (Gibb, 2019, p. 119, n. 1).

is better, what n_0 refers to cannot be a simple event. Rather, n_0 is a causal *process* involving multiple neurons and the constellation of powers of their constitutive parts mutually manifesting, polygenically producing a pleiotropic outcome involving many effects (Molnar, 2003, p. 194). The process exhibits a kind of organic unity, with the powers of the neurons and their parts that are manifested being directed at the final outcome that includes not only n_1 and m_1, but also m_2 as effects. The emergent powers that are constitutive of the desires that are m_1 and m_2 inherit their relative strength from the process,n_0, with m_1 being stronger than m_2. We assume that n_1 would involve a change in the powers of neurons in the prefrontal cortex that would correspond to the strength of m_1 and m_2 and that the process of getting from n_1 to b_1 would involve the mutual manifestation of powers of the network of neurons constitutive ofn_1 mutually manifesting to produce changes in the motor cortex. And the manifestations of the various powers constitutive of the relevant prefrontal and motor regions are picked out by 'n_2'. The various manifestations result in the production of b_1.

How does the foregoing filling in of Gibb's account help us to respond to Davies' argument? We will focus our attention on what is going on with the emergent mental powers. But some more remarks about how powers behave in causal processes is in order.

Given that powers are directed at manifestations with other powers that serve as reciprocal manifestation partners, we can describe powers as "for" or "about" manifestations with other powers. Some of the powers are manifested by being masked by other powers, masking other powers, or by having an additive or subtractive effect on the strength of the manifestation of other powers.

Consider p and q, where q has a subtractive effect on the strength of a manifestation of p, but q does not completely mask p. Suppose that p and q are the postsynaptic potentials of an excitatory synapse and an inhibitory synapse, respectively. They are nearby inputs into the same cell active at the same time. Suppose that p is greater than q. The membrane potential is the algebraic sum of p and q. p and q are each directed at exciting or inhibiting a neuron's firing, respectively. But this is only one of the manifestations at which p and q are directed. p and q are also directed at summing, with the strength of each having a subtractive effect on the other. The outcome of the process is the final strength of the membrane potential. The contributions of p and q are not equivalent. Suppose that if we had the same synapses activated on another occasion and a third excitatory synapse, r, is active, the strength of p (and, hence, the degree of its causal contribution) would be greater and q's strength relative to p and r would be diminished. Importantly, we have an explanation of why each causal element contributes to the degree it does owing to how the elements are synchronically interacting with one another to produce an outcome.

Returning to m_1 and m_2, their relative strength was fixed by the powers manifested in n_0. If the power constitutive of m_1 is a double-preventer, then the strength of m_1 is strong enough to mask m_2's powers directed at producing b_2. Being masked and masking are both manifestations of powers at which they are directed. Ergo, when m_1

masks m_2, they are mutually manifesting, with the outcome being an absence. But it is a causal process.[10] This is important, and we will return to this below.

How does the foregoing help us respond to Davies? Davies is right that, on Gibb's theory, it is metaphysically possible that m_2 violates *Closure* and is part of the cause of b_2 (presumably, this would happen if m_2 interacted with the constellation of powers constitutive of a network of neurons directed at manifesting with the constitutive power of m_2 to produce b_2). But it will not. This is because the constitutive power of the mental state m_2 is not only relatively weaker than that of m_1, but it is weak enough to be masked by m_1. We can add that m_2 will not violate *Closure* because there are no reciprocal manifestation partners in the vicinity that would, together with the constitutive power of m_2, produce b_2. Hence, the potential for m_2 to violate *Closure* would never be actualized. If m_2 were to be rendered causally relevant to some action, it would have greater relative strength vis-à-vis m_1, being sufficient to mask m_1, enabling b_2 to happen. Of course, if this were the case, then things would be different at the neural level as well.

In summary, the laws are what they are owing to the powers of objects and the manifestations at which they are directed. No power will causally trump another power that is strong enough to mask it. Hence, the harmony is owing to the intrinsic properties of agents and how they are situated with respect to one another at the time of exercising intentional agency.

We expect that Gibb will not object to any of the foregoing given that we have articulated widely held commitments held by powers realists about how powers work in causal processes. In what follows, we will be entering less familiar waters and we cannot say for certain who will accompany us, including Gibb.

5 The locality condition: is double-prevention enough?

Suppose that, owing to fleshing out the metaphysics of causation presupposed by Gibb, her theory of mental causation is left unscathed by Davies' *Closure* objection. Not all is well. One worry comes from a fellow causal realist, Jaegwon Kim. Kim has argued that what we are interested in when we are debating mental causation needs to be something more than counterfactual dependence. We need a "'thick' variety of causation in thinking about mental causation" (Kim, 2010, p. 257). So not only do we need to assume causal realism, we need to assume that the sort of causal relevance that obtains when we identify mental causes is causal production. This is because, according to Kim, our concerns about mental causation stem from our caring about human agency, "and agency requires the productive/generative conception of causation" (Ibid., p. 257).

Importantly, causation as production requires satisfying a *locality condition*. Ned Hall characterizes locality as follows: "Causes are connected to their effects via spatiotemporally continuous sequences of causal intermediaries" (2004, p. 225). If putative cases of mental causation must satisfy the locality condition, then the causal

[10] See Buckareff (2018) and Mumford and Anjum (2011) for more on how causal processes can have a zero-sum outcome.

processes involved in exercising intentional agency involve *real* connections between mental causes and bodily behavior (Kim, 2010, p. 257). It is natural at this juncture to ask what real connections are. Real connections, according to Kim, are constituted "by phenomena like energy flow and momentum transfer, an actual movement of some (conserved) physical quantity" (Ibid., p. 257).[11]

Mumford and Anjum (2009) have argued that cases of double-prevention fail to satisfy the locality condition. The reason for this is that double-prevention is not production. On a powers theory of causation, the outcome of a causal process is "the gaining of a new cluster of powers" (Mumford & Anjum, 2009, p. 283). For instance, "Fragility manifests itself in breaking and being broken... [which] means having a new set of powers" (Ibid). So causation on this view can be described as involving a process of the "passing around of powers" (Ibid., pp. 286–287). The idea here is quite simple. The outcome of a causal process involves the inheritance of a power from the manifesting powers constitutive of a cause. Elsewhere, Mumford writes as follows regarding causation as the passing around of powers.

> When a power manifests itself, it sometimes passes on the very same power, as for instance when momentum is passed on from one thing to another, as in Hume's classic case of billiard balls colliding. But in many cases, it is a different power that is passed on. A fragile glass manifests itself in breaking, for instance, and the pieces now have a power to cut which they didn't have when they were a part of the glass. But those pieces, even taken all together, no longer have the power to hold water. (Mumford, 2009, p. 98)

The powers-theory of causation (in particular, as an account of causal production) can thus be regarded as at least a close relative to the sort of conserved quantity view assumed by Kim that was mentioned above.

The powers-theory of causation assumes that genuine cases of causal production satisfy the locality condition.[12] "The passing of powers requires, where it is immediate, spatiotemporal proximity. Cause and effect are to be understood as power and

[11] Incidentally, Gibb (2010) offers a critique of the role afforded conserved quantity views of causation in defenses of physicalism. Specifically, she argues that, while they may capture something that is correct about *physical* causation, the move to saying that a transference of a conserved physical quantity is characteristic of all causal relations is question-begging (see Ibid., pp. 376–377).

[12] A referee for this journal raised worries about whether the locality condition is a commitment of a powers-theory of causation. They offered the example of gravity, in particular where gravitational force is exerted across a vast distance (e.g., between the Sun and the Earth). The worry is that such a view requires that one take such objects to be connected by something such as interacting fields or spacetime regions. On such a view, for instance, the constitutive powers of the gravitational field of the Sun are manifested when partnered with the constitutive powers of the Earth's gravitational field. The worry may be that this renders a metaphysical theory of causation hostage to the sciences. Suppose we learn tomorrow that putative cases of action at a distance do, *in fact*, violate the locality condition. In that case, then we may either need to abandon causal realism entirely or treat the locality condition as typically satisfied but as a non-essential feature of processes of causal production. The second option may lead one to endorse a version of causal pluralism (see, e.g., Cartwright, 2007) on which the locality condition is satisfied with certain processes and not with others. In any case, for our purposes here, the sorts of causal processes we are interested in are ones that would satisfy the locality condition (see Gibb, 2013, p. 200). Such causal processes involve the polygenic production of an outcome by a constellation of powers whose manifestations depend upon their interacting in a way that would satisfy the locality condition (even if it is best to endorse a capacious notion of "local," see Williams, 2019, pp. 144–146).

manifestation where one merges into another in a continuous process" (Mumford & Anjum, 2009, p. 287). Mumford and Anjum insist that double-prevention is not production. No outcome is generated in cases of double-prevention since no powers are passed. Notice that this is the case with Gibb's theory since what happens at the mental level is spatiotemporally disconnected from what happens at the physical level. Mumford and Anjum summarize the problem thusly: "Double-prevention concerns the non-exercise of powers: twice over. A power is prevented from exercising when another also fails to exercise. We have, therefore, two failures of causation. Just as two wrongs do not make a right, two failures of causation do not make a cause" (Ibid., p. 287). If Mumford and Anjum are correct, then Gibb's theory of mental causation does not deliver the goods that Kim insists we are after in a viable theory of mental causation.

This strikes us as being a much more worrisome objection. But the objection rests on the assumption that m_1's role as a double-preventer is not itself part of a causal process that involves the manifestation of powers. Moreover, someone who offers this objection is failing to acknowledge how the process of m_1's double-preventing is part of a larger causal process, the elements of which together can be truthfully described as directed at the agent's performing b_1.

Regarding double-prevention, recall that the constitutive powers of m_1 and m_2 are reciprocal manifestation partners. They are both directed at specific manifestations with one another: m_1 is directed at masking m_2 and m_2 is directed at being masked by m_1. When they are partnered, m_2 is masked, and hence blocked from contributing to the production of b_2, which would violate *Closure*. Importantly, in the case of double-prevention, we have a causal process. Only, in this case, the causal process has a zero-sum outcome, with both the enabling desire m_1 masking m_2, and m_2 being manifested in being masked, respectively. So no intervention occurs, allowing the neurophysiological cause, n_2, to be efficacious.

Notice that in the case just described we have a causal process that satisfies the locality condition. The powers manifested in the causal process are interacting as they would in a causal process with an outcome produced. We have a causal process without causal production. There is a causal process because we have the manifestations of the constitutive causal powers of the desires, m_1 and m_2. But there is no causal production because they effectively cancel each other out. Counterfactually, if the masked desire, m_2, had been considerably stronger, then there would be an outcome generated in the form of a physical intervention. Thus, we have causal relevance: Because if it were not the case that the enabling desire m_1 were manifested, then m_2 would have been effective in blocking the neurophysiological process.

The locality condition is satisfied in a causal process such as we have with the case of double-prevention. But, while satisfying the locality condition is necessary for mental causation, is it sufficient? If a desideratum of a theory of mental causation is that it should suffice to account for our conception of ourselves as agents who produce outcomes in the world, does mental causation as double-prevention do the trick? How we answer this question depends upon how we individuate causal processes. Above, we described each of the atoms in Fig. 1 as corresponding to causal processes involving the activation of constellations of causal powers that are reciprocal mutual manifestation partners. But the boundaries of causal processes extend beyond the activity of powers

whose manifestations directly contribute to the production of the outcome. The activity of any powers that are causally relevant to the production of an outcome are no less a part of the larger causal process. This includes both those powers whose manifestations directly contribute to generating the outcome *and* those whose manifestations *enable* and *structure* the production of the outcome. The entire constellation of both the powers that causally enable the causal production of the outcome and the powers that produce the outcome can together be truthfully described as being directed at a common *telos*. The outcome would not occur without the contributions of both families of powers.

Assuming a view on which an agent is a functionally integrated system of simple objects and their powers, then the agent can be truthfully described as a hybrid complex object with both physical properties and emergent psychological properties. The causal process represented by Fig. 1 is the agent's exercising agency. The outcome at which the collective manifestations of the myriad powers are directed is the bodily movement, b_1. It is their collective activity that causally produces b_1. The contributions of all the powers are not uniform. Some play a productive role. Others play an enabling role. Moreover, there may be other roles for various powers to play. But absent the activity of any of them, we would have a different outcome.

If we are right, then it seems that the threat posed by the causal production objection is a chimera. The account of mental causation as double-prevention as sketched can satisfy the locality problem and provides us with the truthmakers to describe the agent as producing their action.

6 Exclusion worries and going where Gibb may be reluctant to go

In the previous two sections, we have followed Gibb in focusing on the causal role of emergent mental properties as double-preventers. In particular, we have addressed whether the account has the resources to both preserve *Closure* and address the locality problem, providing the truthmakers for representing ourselves as agents who produce changes in the world. Suppose that our additions to Gibb's theory render it immune to the aforementioned worries. There is still another problem. Davies briefly raises an additional problem related to *Exclusion.* He presents a case where we have m_1 but we do not have m_2. In such a case, there is no conflicting desire for m_1 to mask. In such cases, Gibb holds that "there is nothing that my desire must do to permit the causal relation between the relevant physical events to take place" (2013, p. 212). We agree with Davies that this is not very satisfying, but for reasons different from those given by Davies.

Davies argues that Gibb has provided no reason for thinking that the powers constitutive of the desire m_1, would not be manifested in the absence of an opposing desire, m_2. Davies gives the example of an agent having a desire to drink water that accords with the agent's corresponding neurophysiological state. In this case the agent drinks water from a fountain, being caused to do so by both the desire and the neurophysiological state (Davies, 2016, p. 41). This violates *Exclusion* given that the neurophysiological state is assumed to be sufficient for the bodily movement, thus excluding any causal role for the desire. He writes:

> In the absence of a countermanding double-preventing mental event the dispositions of the desire will be manifested, and the thirst quenched. It seems clear then that my drinking from the fountain is overdetermined by events in the physical causal chain and by my desire. Furthermore this is a form of systematic overdetermination: whenever a desire is in accordance with what the physical causal chain delivers, and is not countermanded by other double-preventing desires, it will manifest its disposition to cause that same physical outcome. The Double Prevention model, far from ruling out systematic overdetermination, presupposes it in all those happy cases where the desired outcome is that delivered by the causal chain of events. (Ibid., p. 42)

Contra Davies, a story can be told about such cases on which mental properties are both causally relevant and there is no violation of *Exclusion.* There are two complementary possibilities we will consider. First, the power constitutive of a mental state is causally relevant in virtue of a counterfactual for which it provides a truthmaker. The second possibility would be one that affords a more substantive role to the mental powers in the etiology of outcomes (which does not preclude the counterfactual role for the desire). In particular, the second possibility involves the imparting of form or structure to bodily activities. The first option, taken by itself, while it answers Davies' worry, is not very satisfying for reasons we will articulate. The second option strikes us as more promising. We will expand on these possibilities in turn.

6.1 Mental properties as potential double-preventers

First, consider cases where there is a single desire and no conflicting desire. The power constitutive of the agent's desire to quench their thirst is directed at masking any contrary desire that is equal or weaker in strength. Thus, the power is a counterfactual double-preventer. Given that the neural activity is producing the desired movement, there would be no work for the desire to do. But it is still causally relevant because of its role as a counterfactual double-preventer that, while not a producing cause or an enabling cause, figures in a comprehensive causal explanation of the final outcome. Specifically, the desire would provide the truthmaker for the claim that if there were a contrary desire, then the agent would still get a drink of water.

The foregoing seems like it would answer Davies' worry. But we do not think it is the best response to Davies' objection. The reason why is that Davies' description of the situation is superficial. (For that matter, we are not convinced that Gibb provides a satisfying description of what would occur in this sort of case.) Davies leaves out the other mental states and their constitutive powers that would figure in a psychological explanation of an agent's acting to satisfy their desire for water. Specifically, he leaves out both the fact that the agent has a feeling of thirst and a belief about the presence of a water fountain that can quench their thirst. Both are mental states. The powers constitutive of the feeling of thirst and the belief that there is a water fountain are suitable manifestation partners for the desire to generate an outcome. How can they cause an outcome without violating *Exclusion*?

We assume that epiphenomenalism is off the table as an option. Hence, we shall assume an understanding of causal relevance at the mental level consistent with what

Gibb presents in her account of mental causes as double-preventers. But we shall argue that the role of mental causes need not be limited to double-prevention.

6.2 Mental properties as structuring causes

Following E. J. Lowe, we assume that psychological causal processes and neurophysiological causal processes are complementary (Lowe, 2006, p. 16). That said, the position we sketch in what follows does not parrot the views put forward by Lowe, given that we depart from Lowe on some key points.[13] We will argue that the best way forward is to take mental causes and neurophysiological causes to be directed at different, complementary, outcomes. Specifically, the constitutive powers of neural states (henceforth, 'neuropowers') are directed at bodily movements *qua* types of bodily movements and not at bodily movements *qua* types of intentional actions. The powers constitutive of mental states (henceforth, 'psychpowers') are directed at bodily movements *qua* types of intentional actions. There is no fine-grained one-to-one mapping from types of neuropowers to psychpowers. But there is a coarse-grained mapping. In the case of bodily movements, multiple discrete types of collections of neuropowers with different constituents could provide the emergence base for the same type of psychpower (e.g., there are multiple ways for one to intentionally kick a ball). Neurophysiological and kinaesthetic differences between bodily movements that occur when one is kicking may not affect the identity of what occurs as an intentional action of kicking (again, there are multiple ways to kick the same ball to execute one's intention to score a goal). Borrowing from Lowe, the complete explanation of what occurs requires the irreducible activity of powers at the psychological level and the neurophysiological level. The activation of each type of power—viz., psychpowers and neuropowers—provides the basis for a different type of causal explanation from the other type of power (Lowe, 2008, p. 102). But, *contra* Lowe, we do not hold that mental causation *qua* intentional causation is directed at a specific outcome while physical causation is "blind," not being directed at outcomes in the same way as mental causes (Ibid., p. 110). We expect Gibb would agree with us, given her endorsement of C. B. Martin's metaphysics of powers and causation (Martin takes all causal powers to be directed at specific outcomes with various reciprocal manifestation partners—see Martin (2008, pp. 1–6).

Where we follow Lowe, and think Gibb ought to as well, is in affirming Lowe's claim that both mental causation and physical causation need to be invoked for a full explanation of intentional action (Lowe, 2008, p. 110). Lowe takes mental causes to be "causally responsible for the *fact* that a bodily movement *of a certain kind* occurs, whereas a neural event or set of neural events, is causally responsible for a *particular* bodily movement, which is a particular *event*" (Ibid., p. 110). That the two types of causes are directed at *facts* versus *events* strikes us as based on a rather unclear distinction between facts and events that may fail to carve out any real difference

[13] For one, we do not accept the distinction between the agent/person and their body that Lowe endorses. Rather, as noted, we assume that an agent is a functionally integrated hybrid system composed of simple objects and their properties, including emergent psychological properties.

in the world.[14] So our own theory will not include such a distinction as one of its ontological commitments.

On the view we favor, exercising intentional agency may best be understood as a *hylomorphic* process. In brief, we take it that neuropowers produce a bodily movement which is given a *structure* or *form* by psychpowers. That we have a *bodily process* that is a movement is an outcome of the activation of *neuropowers* and other powers of objects in the agent's environment that are reciprocal manifestation partners. That it is true that what occurs is an exercise of *intentional agency* is owing to the activity of the agent's *psychpowers* and the powers of certain other objects in their environment that are reciprocal manifestations partners. This view is similar to Lowe's insofar as what kind of state of affairs we have (specifically, whether or not it is the tokening of an action-type) when someone's body moves depends upon the activation of their psychpowers. Similar strategies are adopted by Fred Dretske, who takes motivational states to be structuring causes that shape the process that is some particular type of behavior (1988, pp. 42 and 114), and William Jaworski, who has proposed that intentional actions are "structured activities" that result from "the structured manifestations of the powers of an individual's parts and surrounding materials" (2016, p. 177).[15]

Regarding structuring, we assume it is a species of one of Aristotle's four-causes. *Structuring* is *formal causation*. We assume, following Aristotle, that all causes are invoked to answer the question "*Why?"* (Aristotle, *Physics* 194b16-20 (1996)). Formal causes help us explain what type of thing an object, state of affairs, event, process, etc., is. For instance, given its shape, a quantity of bronze is a statue. But formal causes are also invoked when explaining how something manifests its function. Consider the shape and elasticity of a basketball. These properties are powers in virtue of which the ball can do various things. Its shape is a formal cause of the ball's rolling and its elasticity is a formal cause of its bouncing when dribbled. We invoke these properties in answering why the ball rolls and bounces. The structure of the basketball enables one to roll it or dribble it and it also imposes a structure on what the ball can and cannot do.

It is reasonable to ask about the nature of the structuring and formal causal role of psychpowers. While there are important and non-trivial differences between our theory of mental causation and Dretske's, some of Dretske's remarks about structuring causes apply to how we are thinking about them in this article. Dretske distinguishes between *triggering causes* and *structuring causes.* As we understand him, this distinction corresponds to the distinction between *efficient causes* and *formal causes* in Aristotle's metaphysics of causation. He writes that a triggering cause "causes the process to occur *now*," while a structuring cause is "responsible for its being *this process,* one having *M* as its product that occurs now" (1988, p. 42). Elsewhere, he writes that,

[14] Alexander Carruth and Sophie Gibb argue that Lowe's "distinction between fact and event causation is difficult to maintain within most common ontological systems" (2018, p. 150).

[15] None of this is meant to imply that these three philosophers endorse the same metaphysics of mind. They are quite different. Lowe defended a non-Cartesian substance dualism. Dretske endorsed physicalism. Finally, Jaworski endorses a stronger version of hylomorphism than we are presenting here on which psychpowers exert downward causal influence. Limitations of space do not allow us to examine any differences and similarities between what we are proposing and the various proposals offered by Dretske, Jaworski, and Lowe.

"[w]hat thoughts, desires, and feelings explain is not why your arm moves (when you move it intentionally), but why *you* move your arm" (1997, p. 152, emphasis added). This may be as far as the similarities go between our account and Dretske's. But an important take-away is that, on our account (as with Dretske's), the constitutive powers of the representational content of an agent's mental state *qua* intentional state causally explain and structure what some mental state or an action is *about* or *for*. For instance, that an agent's movement can be truthfully described as an intentional punching of another person with whom they are angry is owing to the structuring role of the psychpowers of the agent whose movement we are describing. In brief, psychpowers as structuring causes provide the truthmakers for intentional descriptions of behaviors. Absent the activity of psychpowers, there is no intentional agency, only mere activity or behavior.

Before moving on, it may help to contrast our proposal with one that appears to some, at first glance, to be quite similar, but, in fact, is not.[16] Frank Jackson and Philip Pettit have argued that properties can be causally relevant either by "being efficacious in the production of whatever is in question, or through programming for the presence of an efficacious property" (1990b, p. 115). These correspond, accordingly, to two types of causal explanations, namely, *process explanations* and *program explanations* (Jackson & Pettit, 1988, p. 388). A property does not program for the presence of an efficacious property by figuring in the productive process. Rather, its programming role is limited to ensuring the presence of a "suitably efficacious" property that is required for the tokening of a type of causal process (Jackson & Pettit, 1990b, p. 114). Importantly, on this view, properties that figure in program explanations, such as mental properties on Jackson and Pettit's view, "do not do any work in producing [an outcome]," they are "perfectly inert." Their role is limited to "ensuring that there would be some property there to exercise the efficacy required" (ibid., p. 114). While Jackson and Pettit agree with us and Gibb that mental properties are dispositional properties, they take dispositional properties to be causally impotent (1990a, p. 203).[17] On the other hand, we take them to be causally *potent* (they are, after all, causal *powers*)! They do stuff. The fact that we assume powers realism alone is a significant difference between a theory of mental causation like ours and Jackson and Pettit's program explanation account. On our view, what mental properties do is not the same as what the constitutive powers of the basal conditions do in the causal processes, but the mental properties still play a causal role in the causal process that is picked out by a process explanation.

We can now tell a story of what goes on in cases where there is no actual double-prevention that occurs. The agent's reciprocal psychpowers constitutive of their desire for water, their feeling of thirst, and their belief that there is a water fountain are emergent properties whose mutual manifestation polygenically causally structures what is happening as a process of thinking about what to do. An emergent psychpower of intending (the structure of which is the result of the powers that were manifested in the process of practical thinking) to drink water partners with the agent's psychpowers

[16] A referee for this journal suggested that there are similarities between our account and the program explanation account of Jackson and Pettit.

[17] See Prior et al. (1982) for more details about the view of dispositional properties that is assumed by Jackson and Pettit. For a critique, see Heil (2003, pp. 87–89).

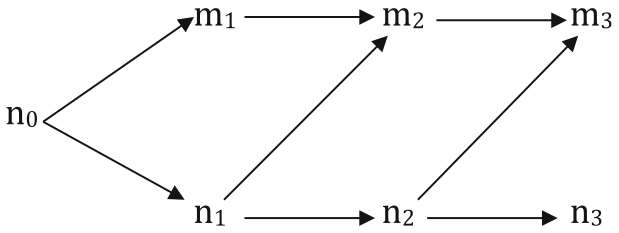

Fig. 2 Mental causation as structuring

constitutive of their perceiving the location of the water fountain, resulting in the imposition of a structure on the activity of moving toward the water fountain and taking a drink, making it true that the agent acts intentionally. The entire process can be described as involving the activation of powers converging on a common *telos*, in this case, the drinking of water.

The account we are presenting follows the theory offered by Gibb to the extent that we agree that psychpowers enable the occurrence of what is causally produced by the neuropowers. The difference is that we are taking the enabling work of psychpowers to extend beyond the role Gibb assigns them as double-preventers. Psychpowers structure what is being produced by the neuropowers, providing the truthmakers for representing what occurs as an intentional action as opposed to, say, a reflexive behavior. The main point of departure from Gibb lies in our taking the enabling/structuring work done by psychpowers to not being limited to the role of being actual or counterfactual double-preventers. The psychpowers interact with one another and confer structure on the outcomes of neuropowers.

How does the foregoing fit within a complete neo-Aristotelian framework invoking four-causes?[18] On our account, the fundamental objects constitutive of the agent *qua* organism and their environment are the *material cause* of the agent's action. The neuropowers of the agent are the *efficient cause*. The psychpowers of the agent are the *formal cause* of the action. The *final cause* is the *telos* at which the combined activated powers in the causal process are polygenically directed. None of these elements duplicates the work of the other elements. Together, they provide us with the resources to give a comprehensive causal explanation of an agent's exercise of intentional agency. Importantly, the manifestations of the psychpowers as the formal cause of the action do not have a directly productive role in the production of some bodily activity[19] (their productive role is indirect as part of the total cause of what we can truthfully describe as an intentional action) and their activation does not violate closure given that their work is done exclusively at the mental level (we will say more about this point and how this relates to double-prevention in a moment).

What would the foregoing look like? Return to the case of the agent desiring to slake their thirst who believes that getting water from the water fountain will satisfy their desire. Consider the following neuron diagram (Fig. 2).

[18] See Aristotle, *Physics* 194b23-35 (1996). We are not suggesting that the way of carving up the four causes we present actually tracks how Aristotle treated his four causes.

[19] Note that we assume that bodily activity is not limited to overt actions. Mental actions involve bodily activity in the form of neurophysiological activity.

The desire and belief, m_1, and the corresponding neural state, n_1, have a common neurophysiological cause, viz., the manifestations of the constitutive neuropowers of n_0. The constituent neuropowers of n_1 interact and have a pleiotropic causal outcome, n_2 (say, signals to the motor region of the brain) and m_2 (an intention to act). The constituent psychpowers of the desire and belief at m_1 are reciprocal manifestation partners and have as an outcome the conferral of structure on m_2 as a proximal intention that represents a specific action-plan to get a glass of water (where the plan is inherited from the belief and desire). The manifesting constitutive neuropowers of n_2 (e.g., the activation of the agent's motor cortex) has as its pleiotropic outcome a bodily movement n_3 that is purposeful behavior, m_3. The constitutive psychpowers of m_2 would partner with the constitutive psychpowers of the agent's awareness of their immediate environment, with the outcome of their manifestations being an intentional structure conferred upon m_3 as a purposeful behavior with a specific goal that is inherited from the agent's intention. That is, owing to constitutive powers of m_2, what we have with the occurrence of m_3 and n_3 is purposeful behavior that is the tokening of a particular action-type.

This is all very quick and there are no doubt questions about some of the details of the account on offer. But this is just a sketch. Considerations of space alone will not allow us to develop this further in this article. That said, we believe this sort of hylomorphic theory of intentional agency and the role of emergent mental causes in the process is one that we think is worth exploring and articulating in more detail.

Returning to the worry expressed by Davies, on the account on offer, we do not have a case of overdetermination at the neurophysiological level. The constellations of manifesting reciprocal neuropowers are sufficient for their outcome. But the activity of the neuropowers does not exclude the activity of the psychpowers because they are directed at different things and play distinct causal roles. The powers at each level have distinctive causal and explanatory functions with respect to understanding an agent's actions and neither is reducible to the other, even if the emergent psychpowers are ontologically dependent upon the emergence base consisting of manifesting neuropowers. The neuropowers are directed at generating bodily movements of an organism while the psychpowers are directed at structuring the movements as intentional actions. Their roles are complementary and not in competition with one another. Together, they are ultimately directed at a common *telos*, which is the goal of the agent in exercising intentional agency.

One final implication of this model is worth mentioning. Our theory takes the double-preventive role of psychpowers to involve the psychpowers constitutive of a desire for an outcome, m_{n}, preventing a desire for $\sim m_{\mathrm{n}}$ from preventing the token action's having the structure of *A*-ing in order to bring it about that m_{n}. All of the causal work is on the mental level, having to do with the intentional structure of what occurs. The desire for $\sim m_{\mathrm{n}}$ will not exert downward causal influence. It is not metaphysically possible that *Closure* will be violated. Rather, what is possible is that the token action is of a different type, directed at a different outcome. But that would depend upon the basal conditions being such as to render the desire for $\sim m_{\mathrm{n}}$ being stronger than the desire for m_{n}. Thus, the psychpowers constitutive of the desire for m_{n} that play a double-preventive role enable what occurs to be an action of type *A* rather than another action type or a basic omission.

7 Pulling it all together

The foregoing can now all be put together to provide a more detailed metaphysic of mental causation than Gibb delivers. We take the account to build on Gibb's account. But the final story takes us quite some distance from the actual proposal she offers.

We have offered reasons for thinking that once we expand on and clarify the ontological commitments of Gibb's theory of mental causation, it can escape unscathed from Davies' worries related to *Closure.* With some minor modifications (specifically, shifting to a framework that focuses on causal processes rather than causal chains of events) her account has the resources to also respond to worries related to the locality condition for causal realism. But, given that Gibb's account only mentions the double-preventing role of desires and then assigns psychpowers no role in cases where there is no countermanding psychpower, her account runs into the *Exclusion* worries articulated by Davies discussed above. We have argued that the strategy of desires simply playing no role in the causation and causal explanation of an agent's action when they are not double-preventers seems like an inadvisable path to take for two reasons. First, a desire's constitutive psychpower is still poised to play the role of a counterfactual double-preventer. But, more importantly, there is still an important structuring/formal causal role left for a desire in the absence of a contrary desire given that there are other relevant psychpowers possessed by an agent that are reciprocal manifestation partners with which the constitutive psychpowers of the desire would interact.

What we have now is a theory of mental causation for the emergent property dualist that builds upon the work of Gibb, Lowe, and others, but departs from them in ways discussed above. The account can be briefly summarized as follows. In cases of an agent acting, strongly emergent mental properties are caused by an emergence base consisting of neural properties. Both the properties in the emergence base and the emergent properties are causal powers–which we have referred to as 'neuropowers' and 'psychpowers', respectively. The neuropowers also cause an outcome consisting of a corresponding constellation of neuropowers that are together directed at a bodily movement. The constellation of psychpowers are together directed at a specific action-type. In cases where there is an opposing desire directed at a different action-type, the total weight of opposing powers mask the power, functioning as double-preventers, enabling the neuropowers in the emergence base to produce the bodily movement at which they are directed to be truthfully described as a tokening of a particular action-type. But the role of the psychpowers does not end with double-prevention and it is not limited to this role. Their role as enablers extends to structuring what occurs as being an intentional action and not just mere behavior. The psychpowers give form to the bodily behavior, providing the truthmakers for accurately representing it as an exercise of intentional agency.

Again, we are now some distance from Gibb's proposal. But the account we have offered still follows the spirit of Gibb's theory but edges the property dualist closer to an account of mental causation like Lowe's. We think it is a promising way forward for those attracted to emergent property dualism who also accept a version of causal realism based on an ontology of causal powers.

8 Conclusion

The goal of this article has been to expand and elaborate Gibb's theory of mental causation as double-prevention, trying to fill in some gaps to render it immune to certain objections in the literature. In doing so, we have acknowledged that in order for her account to avoid some objections and be consistent with a fuller picture of what our mental lives are like, then we must expand the theory beyond mental causation as just double-prevention. That being said, we believe the additions we have made to Gibb's account are implied by the ontological commitments of her theory. The theory of mental causation that results is not Gibb's own theory, but it is consistent with the central claims of her theory, and, more importantly, it is more consistent with the declared ontological commitments of her theory of mental causation. By fleshing out these commitments, we believe that the resultant theory of mental causation is more comprehensive and is better suited to provide the truthmakers for purposive explanations of intentional actions. We are not prepared to declare the account a success. Moreover, we are uncertain whether Gibb would accept the foregoing modifications to the account. She may deny that they are implied by any of the stated ontological commitments of her theory. Our hope is that the fixes to the theory we have provided will prove to be sufficiently useful to warrant acceptance by those who wish to defend dualism by adopting a strategy similar to Gibb's.[20]

Declarations

Conflict of interest The authors have no known conflicts of interest in submitting this article for review.

References

Aguilar, J., & Buckareff, A. (2022). Agency and causation. In L. Ferrero (Ed.), *Routledge handbook of philosophy of agency* (pp. 27–36). Routledge.

Aristotle. (1996). Physics. In R. Waterfield (Ed.), *Oxford World's classics: Aristotle: Physics.* Oxford University Press.

Buckareff, A. (2018). I'm just sitting around doing nothing: On exercising intentional agency in omitting to act. *Synthese, 195*, 4617–4635.

Buckareff, A. (2022a). Disentangling dispositions from powers. *Dialogue: Canadian Philosophical Review, 61*, 107–121.

Buckareff, A. (2022b). Two-way powers as derivative powers. In M. Brent & L. Titus (Eds.), *Mental action and the conscious mind* (pp. 228–254). Routledge.

Carruth, A., & Gibb, S. (2018). The ontology of E. J. Lowe's substance dualism. In A. Carruth, S. Gibb, & J. Heil (Eds.), *Ontology, modality, and mind: Themes from the metaphysics of E. J. Lowe.* Oxford University Press.

Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge University Press.

Davies, K. (2016). Powers, double prevention, and mental causation. *Metaphysica, 17*, 37–42.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. The MIT Press.

[20] The authors wish to thank two anonymous referees for this journal for their very helpful feedback on an earlier version of this article. Some of Andrei Buckareff's work on this article was generously supported by a grant from the John Templeton Foundation (Grant #62683, "Panpsychism and Pan(en)theism: Philosophy of Mind Meets Philosophy of Religion). None of the views presented in this article necessarily represent those of the John Templeton Foundation.

Dretske, F. (1997). *Naturalizing the mind*. The MIT Press.
Gibb, S. (2010). Closure principles and the laws of conservation of energy and momentum. *Dialectica, 64*, 363–384.
Gibb, S. (2013). Mental causation and double prevention. In S. Gibb, E. J. Lowe, & R. Ingthorsson (Eds.), *Mental causation and ontology* (pp. 193–214). Oxford University Press.
Gibb, S. (2015a). Defending dualism. *Proceedings of the Aristotelian Society, 115*, 131–146.
Gibb, S. (2015b). The causal closure principle. *Philosophical Quarterly, 65*, 626–647.
Gibb, S. (2015c). Physical determinability. *Humana Mente: Journal of Philosophical Studies, 29*, 69–90.
Gibb, S. (2017). The mental causation debate and *qua* problems. In M. P. Paoletti & F. Orilia (Eds.), *Philosophical and scientific perspectives on downward causation.* Routledge.
Gibb, S. (2019). Causal closure. In S. Gibb, R. Hendry, & T. Lancaster (Eds.), *The Routledge handbook of emergence* (pp. 111–120). Routledge.
Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 225–276). The MIT Press.
Heil, J. (2003). *From an ontological point of view*. Oxford University Press.
Jackson, F., & Pettit, P. (1988). Functionalism and broad content. *Mind, 97*, 381–400.
Jackson, F., & Pettit, P. (1990a). Causation in the philosophy of mind. *Philosophy and Phenoenological Research, 50*, 195–214.
Jackson, F., & Pettit, P. (1990b). Program explanation: A general perspective. *Analysis, 50*, 107–117.
Jaworski, W. (2016). *Structure and the metaphysics of mind: How hylomorphism solves the mind-body problem*. Oxford University Press.
Kim, J. (1976). Events as property exemplifications. In M. Brand & D. Walton (Eds.), *Action theory* (pp. 159–177). D. Reidel.
Kim, J. (1989). Mechanism, purpose and explanatory exclusion. *Philosophical Perspectives, 3*, 77–108.
Kim, J. (2005). *Physicalism, or something near enough*. Princeton University Press.
Kim, J. (2010). Causation and mental causation. In J. Kim (Ed.), *Essays in metaphysics of mind* (pp. 243–262). Oxford University Press.
Lombard, L. (1979). Events. *Canadian Journal of Philosophy, 9*, 425–460.
Lombard, L. (1998). Ontologies of events. In S. Laurence & C. Macdonald (Eds.), *Contemporary readings in the foundations of metaphysics* (pp. 277–294). Blackwell.
Lowe, E. J. (2000). Causal closure principles and emergentism. *Philosophy, 75*, 571–585.
Lowe, E. J. (2006). Non-Cartesian substance dualism and the problem of mental causation. *Erkenntnis, 65*, 5–23.
Lowe, E. J. (2008). *Personal agency: The metaphysics of mind and action*. Oxford University Press.
Malcolm, N. (1968). Conceivability and mechanism. *The Philosophical Review, 77*, 45–72.
Martin, C. B. (2008). *The mind in nature*. Oxford University Press.
Molnar, G. (2003). *Powers: a study in metaphysics*. Oxford University Press.
Mumford, S. (2009). Passing powers around. *The Monist, 92*, 94–111.
Mumford, S., & Anjum, R. (2009). Double prevention and powers. *Journal of Critical Realism, 8*, 277–293.
Mumford, S., & Anjum, R. (2011). *Getting causes from powers*. Oxford University Press.
Prior, E., Pargetter, R., & Jackson, F. (1982). Three theses about dispositions. *American Philosophical Quarterly, 19*, 251–257.
Schneider, S. (2012). Why property dualists must reject substance physicalism. *Philosophical Studies, 157*, 61–76.
Taylor, H. (2018). Powerful qualities and pure powers. *Philosophical Studies, 175*, 1423–1440.
Williams, N. (2019). *The powers metaphysic*. Oxford University Press.