This is the almost final version of a paper forthcoming in *Inquiry*. The text of the published version may differ slightly.

The Disappearing Agent as an Exclusion Problem

Johannes Himmelreich

The disappearing agent problem is an argument in the metaphysics of agency. Proponents of the agent-causal approach argue that the rival event-causal approach fails to account for the fact that an agent is active. This paper examines an analogy between this disappearing agent problem and the exclusion problem in the metaphysics of mind. I develop the analogy between these two problems and survey existing solutions. I suggest that some solutions that have received significant attention in response to the exclusion problem have seen considerably less attention in response to the disappearing agent problem. For example, one solution to the exclusion problem is to reject the exclusion assumption. Analogously, one solution to the disappearing agent problem could be to deny the claim that the agent-causal approach and the event-causal approach are mutually exclusive. Similarly, proportionality theories of causation, a solution to the exclusion problem, can be transferred to the disappearing agent problem. After establishing the plausibility of the analogy between the two problems, I examine how this latter solution in particular can be transferred from the one problem to the other.

1 Introduction

The metaphysics of agency is the venue for a chasmic rivalry between two competing approaches known as the agent-causal and the event-causal approach.[1] The two approaches agree insofar as both analyze agency in terms of causation, but they differ, among many other things, in what they identify as causes in action. The agent-causal approach says that the relevant causes are agents qua their persisting substance.[2] By contrast, the event-causal

[1] The terminology follows Schlosser (2015). I take this binary division of the conceptual landscape for granted but register a point of dissatisfaction. Not all proponents of the event-causal approach take the relata in causal relations to be events. Some instead identify exercises of agency with interacting causal powers manifested in causal processes (e.g. Stout 1996; 2005, chap. 6; Mumford and Anjum 2011, 204; Buckareff 2012; and, to my reading, Steward 2012a; 2016).

[2] For the sake of ecumenism, as I explain shortly, I henceforth omit the "qua substance" qualification.

approach says that the relevant causes are agents' mental entities. In the eyes of the agent-causal approach, the event-causal approach faces the disappearing agent problem:[3] In analyzing agency reductively in terms of causation by an agent's mental entities, the event-causal approach fails to account for the fact that the agent herself is active.

This paper is on the metaphilosophy of action. Some contend that this disappearing agent problem is analogous to the exclusion problem (Mayr 2011, 10–12; Buckareff 2012, 115; Steward 2012b, 197–247). But whether this contention is sound has not yet been answered. In this paper, I cast the metaphysics of agency into the well-known dialectic of the metaphysics of mind. I argue that the contention is indeed sound, and I detail how the two problems are analogous. I identify important limitations of the analogy and I illustrate benefits of such a theoretical unification, including an example of how new theoretical avenues might open up.

The analogy allows us to transfer solutions that have been put forth in response to one problem and apply them to the other. For example, one solution developed in response to the exclusion problem is to reject the exclusion assumption. Because this is an assumption that both arguments substantively share, this claim immediately transfers over from the exclusion problem to the disappearing agent problem. Moreover, theories of causation that have been developed in response to the exclusion problem, such as proportionate causation and distinctions between structuring and triggering causes or causal relevance and causal efficacy, may provide plausible sketches to build out approaches to agent causation. Of course, transfers in the other direction — from the disappearing agent problem to the exclusion problem — seem also plausible.

The paper has three parts. I first present a version of the disappearing agent problem in section 2. Then I discuss the exclusion problem in section 3. In section 4, I survey solutions on each side of the analogy and identify limits of the analogy. In section 5, I sketch as an example how one solution of the exclusion problem can be transferred to the disappearing agent problem to build an account of agent causation.

The project here is partly procrustean and partly heretical. Every argument has its received dialectic and canonical presentation. The way in which I gloss the rivalry between agent-causal and event-causal approaches may appear defective to those steeped in this dispute. Except for section 5, I only speak of "mental entities" and "results of actions" without specifying their ontological categories. But because ontological categories are front and center in the received dialectic, my project might be considered lacking in

[3] It must be noted that this problem is different from another problem that goes by the same name (Pereboom 2014a, chap. 2; Clarke 2017).

ontological seriousness (Buckareff 2012, 106). Similarly, I uniformly classify as defending an agent-causal approach both those who understand the agent as a substance causing results of actions (e.g. Mayr 2011; Whittle 2016), as well as those who understand the agent as a causal process or a system of causal powers (Buckareff 2012; 2018; Stout 2007; 2012).[4] Stepping back from important details in this way helps to bring out the analogy. I aim for this paper to offer an alternative dialectic and to illustrate, within the constraints of a journal article, the fruitfulness of this alternative picture on the assumption that when a picture is fruitful, it is plausible (Gutting 2009).

This project has been met with resistance from two opposing ends especially concerning the suggestion — a central move in the paper — to adopt a picture of levels in understanding the disappearing agent problem. Some readers have dismissed this as unmotivated and implausible. They see bare assertions and no reason why levels should be understood in any other way than merely descriptively or metaphorically — and particularly not ontologically and applicable to the disappearing agent problem. Others, by contrast, have found that this suggestion to understand the disappearing agent problem in terms of ontological levels is too obvious to warrant publication. They contend that the picture of levels pervades such a great many debates and conflicts that exclusion-style worries crop up all too frequently. For these, the fact that two debates exhibit such exclusion-style worries is therefore insufficient for supporting an analogy claim. So, what is absurd to one is obvious to another. I mention this to support a plea for skepticism: On this issue at hand, the opposite of your conviction may be as obvious to others as is yours to you.

2 The Disappearing Agent

Suppose you want to reach for a cup in front of you on the table. You reach for it, grab it with your hand and, as a result, the cup is in your hand. This result has a cause.[5] Three candidate causes may come to mind. First, there is the individual. You, the agent, are the cause of the cup now being in your hand. This is the claim that proponents of the agent-causal approach emphasize. Among other things, this position finds support in the way

[4] Although Stout (2007; 2012) does not self-identify as a proponent of the agent-causal approach, he can be read as such (Bishop 2007, 158).

[5] I talk of "result of action" to avoid the controversial claim that agents cause their *actions* (Alvarez and Hyman 1998). Furthermore, set aside for now whether "result" refers to a fact, an event, a property, etc. I also remain neutral between so-called product views and component views about the relation between an action and its cause.

actions are described in natural language.[6] The second candidate cause is mental entities, such as the event, state, or fact of having a certain intention. You *wanted* to reach for the cup. The claim that your wanting caused the cup to now be in your hand is emphasized by proponents of the event-causal approach. Third, there are physiological entities: The firing of neurons in your motor cortex and other goings-on in your neurophysiological system cause the cup to be in your hand. Neuroscience explains bodily movements with causal claims like these. Each of these three claims seems plausible, at least in some contexts. How do they fit together?

Each of these claims can be seen as operating on a given level. I say more about how this metaphor of levels should be understood in section 3. For now, entertain the idea at least for illustrative purposes that an agent's action can be seen as involving three levels: the *individual level*, the *intentional level* with the individual's mental entities (you could also call this the psychological level), and the *implementation level*, where the neurophysiological system operates.

There is one way in which this picture of levels can be helpful, already before specifying what levels are. This picture allows you to distinguish approaches about the metaphysics of agency by the level on which they locate the cause of an action's results. Hence, I distinguish the approaches by the causal claims that they stake out. An *agent-causal approach* locates the cause of an action on the individual level: agents are causes. In contrast, an *event-causal approach* locates the cause of an action on the intentional level. This approach claims that for each action there is a mental entity causing the action's results — mental entities are causes.

This way of distinguishing agent-causal and event-causal approaches should not mask the fact that these approaches present two fundamentally different ways of understanding human agency. Distinguishing the approaches by their causal claims is only one way of telling them apart. I say more about an alternative way of distinguishing these approaches shortly. But let me illustrate upfront how deep the differences between the two approaches run. Proponents of the agent-causal approach might contend that proponents of the event-causal approach commit something like a category mistake. They might argue that human agency involves the agent as such. The intentional level is, therefore, just not the right level

[6] In expressions such as "you take the cup," you count as the agent of taking the cup in the linguistic sense (a semantic concept). Some take this to support the claim that you are an agent of taking the cup in the philosophical sense. Chisholm (1976, 199) makes such an argument: "[S]tatements [such] as 'Jones killed his uncle' and 'Jones raised his arm' are sometimes true; ... they imply that there is a certain event or state of affairs that an agent caused or brought about." For similar arguments, see Lowe (2008, 122) and Steward (2012b, 200). However, it should be noted that Lowe (2008) also pursues further additional arguments.

for doing metaphysics of agency. This is the categorical argument against event-causal approaches that does not rest on any idea about exclusion. The disappearing agent problem is a separate and, in some ways, more conciliatory argument against the event-causal approach.[7]

One *locus classicus* of the disappearing agent problem is, when taken by itself, somewhat ambiguous between this categorical argument and the disappearing agent problem. Hornsby (2004, 2) formulates the disappearing agent problem as follows. She refers to the event-causal approach, because it enjoyed greater popularity at the time, as the "standard story" and she writes:

> *[T]he story leaves agents out*. Human beings are ineliminable from any account of their agency, and, in any of its versions, the standard story is not a story of agency at all.

On the event-causal approach, because it takes mental entities to be causes, the agent herself plays no active role. The approach therefore fails to satisfy a crucial desideratum, or so I read Hornsby, to account for an agent's activity. Early proponents of the disappearing agent problem are Melden (1961) and Taylor (1966):

> It is futile to attempt to explain conduct through the causal efficacy of desire — all that can explain is further happenings, not actions performed by agents. The agent confronting the causal nexus in which such happenings occur is a helpless victim of all that occurs in and to him (Melden 1961, 128).

> There must ... be a reference to myself as active being. ... [I]n acting, I make something happen, I cause it, or bring it about. ... It is plain that whatever I am, I am never identical with any such event, process, or state as is usually proposed as the "real cause" of my act, such as some intention or state or willing (Taylor 1966, 111).

Also Goldman (1970, 81) discusses the disappearing agent problem.[8]

> [T]here is the problem of agent-causation versus want-and-belief causation. If the acts of an agent are caused by his wants and beliefs, how can *he*, the agent, be considered their cause? (Goldman 1970, 81).

---

[7] The expression "disappearing agent" seems to be due to Velleman (1992). Recently the disappearing agent problem has been raised by Alvarez and Hyman (1998, 227–28), O'Connor (2005), Lowe (2008, 159–61), Mayr (2011, 6–13), Steward (2012b, 197–247), Buckareff (2012), Pereboom (2014b; 2014a), Brent (2017), and Velleman (2020, 62–63). Aguilar and Buckareff (2010, 12–14) give an overview.

[8] Goldman (1970, 82–83) is not himself a proponent of the disappearing agent problem because he denies the exclusion assumption and holds that the agent as well as their mental entities may cause an act.

Two things should be clear from these passages. First, activity is understood as causation. More precisely, an agent counts as active only if she is a cause. Second, agent-causal and event-causal approaches are taken to be exclusive. An approach can be either agent-causal or event-causal but it cannot be both. Note how Goldman (1970) suggests as much by writing "versus." Taylor (1966, 111) writes that "if ... I sometimes cause something to happen, this would seem to entail that it is *false* that any event, process, or state not identical with myself should be the real cause of it." Similarly, Velleman (2020, 63) writes that causes such as muscle contractions and nerve impulses "crowd me out of being the cause." This exclusion assumption is central to the disappearing agent problem. Otherwise, as far as causation is concerned, there would be no conflict between the two approaches.

To bring out this exclusion assumption with full force, the disappearing agent problem can be stated as a deductive argument with two premises. The first assumption is that agents are causes. That is, for each action there is an agent causing the action's result. This captures the idea that the agent is active. The agent may cause things in virtue of her substance, her powers, or her being a causal process (cf. Whittle 2016; Buckareff 2017; Stout 2007). The second assumption is that this agent-causality, however specified, excludes event-causality. This is the exclusion assumption. Any approach about the metaphysics of agency entails that agents are causes or that events (mental or physiological) are causes, but not both. It follows that any event-causal approach conflicts with the assumption that agents are causes.

Already at this point, the hypothesized analogy with the exclusion problem is instructive. It is well known from the literature on the exclusion problem that the exclusion assumption can be used either in an upward or in a downward direction (List and Menzies 2009, 490). Interestingly, this difference can be seen in some places in the literature on the disappearing agent problem. The proponents of agent causation, such as Hornsby, Melden or Taylor, run the argument in a *downward direction*. They start with the premise that there is causation on a higher level and infer that there cannot be causation on a lower level. But you could reverse this argument and use the exclusion assumption in an *upward direction*. Nagel (1986, 110–11), Velleman (1992; 2020, 62–63), Lowe (2008, 159–61) and Steward (2012b, 197–247),[9] among many others, start with the assumption that physiological entities are causes and that this hence undermines agent-causation from below. This difference in starting assumptions may indicate different underlying concerns among participants in this literature. Moreover, it may bring out why, to some, the debate between agent-causation and event-causation may seem to have reached something like an impasse. If

[9] The assignment is less clear for the latter two. Importantly, they each offer further arguments not discussed here.

both parties tacitly accept the exclusion assumption, their respective starting points are not reconciled easily (but see Clarke 1993; Bishop 2007, 158).

Further versions of the same argument can be distinguished by the pair of levels on which the arguments concentrate. Melden, Taylor and others, see the individual level conflict with the intentional level (e.g. desires and intentions). By contrast, Nagel and Velleman see agency (the individual level) as clashing with the implementation level (e.g. chemical reactions and muscle movements).

Some might worry at this point that I mischaracterized the agent-causal and event-causal approaches. Instead of causal claims, a deeper disagreement between the approaches concerns reduction and fundamentality. The agent-causal approach contends that the individual level is fundamental and that the intentional level ontologically depends on the individual level. This fundamentality claim can concern agents themselves, causation, and agency (as a relation distinct from causation). In this vein, Buckareff (2012, 106) characterizes the agent-causal approach by the claim that "agent-causation is ontologically irreducible" (cf. Chisholm 1976, 199). Should agent-causal and event-causal approaches be distinguished by their respective claims about fundamentality instead of their respective claims about causes?

I concede that the broader conflict between agent-causal and event-causal approaches concerns fundamentality of agents, causation, or agency respectively. Because the presentation here only needs to assume that the causes operate on different levels, but not what the ordering of these levels is, the view that I present is not committed to taking sides in this conflict about fundamentality. You could restate my reconstruction of the disappearing agent problem with the additional assumption that the individual level is more fundamental than the intentional and the implementational level. Hence, I set issues of irreducibility and fundamentality aside to concentrate on the causal claims of the respective approaches. Likewise, Steward (2012b, 203) resists distinguishing the approaches by their respective fundamentality claims on the grounds that this conflates two issues – causation and reduction – that, although not independent, should be kept separate.

Some might also worry that I mischaracterize the disappearing agent problem by seeing it as a problem of agential activity. Recently, the literature discusses a related but different problem under the same name (Pereboom 2014a; 2014a, chap. 2; Clarke 2017). In this other problem, the agent is active in the sense that she settles between consequences in situations of motivational parity. This other problem arises specifically in the context of *libertarian* event-causal approaches. I am instead interested here in the disappearing agent problem understood as competing causal claims irrespective of how causing relates to settling and irrespective of whether this causation makes for free agency.

3 The Exclusion Problem

A well-known problem arises for certain views in philosophy of mind claiming that bodily movements are caused by mental entities (Malcolm 1968; Kim 1989; 1998; 2005; Bennett 2007; List and Menzies 2009). This problem is known as the exclusion problem. But before turning to this argument, the metaphor of levels must be spelled out. I follow an analysis of levels of description detailed elsewhere (List 2019).

Philosophers and scientists often speak of levels of description or explanation considering that many phenomena can be described in different ways. A car crash can be described by Newtonian mechanics, psychology, sociology, or in everyday concepts. The Newtonian mechanics description will describe the car crash in terms of forces and momentum. The psychological description will involve mechanisms of attention or psychological heuristics or biases. The sociological description might reference the driver's age and gender and the level of risk-aversion of his social group. A description in terms of everyday concepts might be that "the driver thought that the road was clear, sped around the corner and crashed into the oncoming car."

To spell out what levels are more formally, a level of description or explanation is represented by two elements (see List 2019). The first element of a level is a language, which can be as simple as propositional logic. This language defines which descriptions are admissible at a given level. Each sentence in a language is a description. The second element of a level is a way of determining which of the admissible descriptions are true. Although this is not without problems, assume that the truth-value of a description is settled by a possible world. Let the set of all worlds relative to a language be the minimally rich set of worlds that are needed to assign a truth-value to any description in the language. In this way, each description is associated with a subset of all possible worlds (again, for details see List 2019). In sum, a level is a pair consisting of a language and the set of worlds that is induced by descriptions in this language.

Already at this point, even if only very roughly, this idea of levels can be linked up with the three-level model of agency above. The individual level, for example, can be associated with our natural language. The intentional level could be associated with describing actions in terms of psychology by making reference to mental phenomena such as cognitive mechanisms or intentions. The implementation level could be associated with, for example, theories of biology or neuroscience.

The exclusion problem arises for theories in philosophy of mind that entail the following two claims. The first claim is that the level of physiological entities determines the level of mental entities. This entails that mental entities are distinct from physiological entities

and there cannot be a change in someone's mental entities without a change in her physiological entities. The second claim is that mental entities cause (or are causally sufficient for) bodily movements. This claim is understood as a claim about causation, not about causal explanation. These two claims — dependence of the mental and causation by the mental — are central clams of *nonreductive physicalism*. But these claims, together with other *prima facie* plausible assumptions, lead to a contradiction when you also accept the following exclusion principle (List and Menzies 2009, 490), which generalizes the exclusion assumption that we encountered earlier in the context of the disappearing agent problem.

Exclusion Principle
For all distinct s and s^*, such that s^* is higher-level than s, s and s^* do not both cause x.

Some such principle is used in different canonical formulations of the exclusion problem (Kim 1989, 44; Bennett 2008, 281; List and Menzies 2009, 478; Kroedel 2015, 362). Sometimes it states that "no effect has more than one sufficient cause unless it is overdetermined" (Bennett 2008, 281) or that "[t]he effects of mental events are not systematically overdetermined" (Kroedel 2015, 362). But the differences in formulation are insubstantial for the purposes here (cf. List and Menzies 2009, 490).

Nonreductive physicalism faces what is known as the exclusion problem in that it seems to imply that physiological entities are not causes. Nonreductive physicalism assumes that mental entities are causes, and by the downward application of the exclusion principle, it follows that physiological entities are not causes.

Exclusion Problem for Nonreductive Physicalism
Physiological entities are causes. For each bodily movement, there is a physiological event that causes it.
Dependence. Mental entities are higher-level than physiological entities and are distinct from them.
Mental entities are causes. There are some bodily movements that are caused by mental entities.
Problem. These three claims are inconsistent with the exclusion principle.

As Bennett (2008, 281) puts it, the exclusion problem consists in the question of "how to avoid commitment to an apparently inconsistent set of propositions." Including the option to give up the exclusion principle, four responses are possible. Each corresponds to giving up one assumption of the argument. Each of these responses has been developed in detail elsewhere, but it is helpful to sketch them out here to identify unexplored solution counterparts across the two problems.

The first response to the exclusion problem is to deny that physiological entities are causes. Proponents of nonreductive physicalism could contend that bodily movements are *not* caused by firings of neurons, muscle movements, and the like. On the resulting position, you maintain the tenets of nonreductive physicalism and accept the exclusion principle. Yet, this position looks radical because many find it obvious that physiological entities are causes. Nevertheless, something can be said to defend this position. The position is supported by *proportionality theories of causation* where the general idea is that physiological entities are too fine-grained, too specific, or too fragile relative to the event of a bodily movement (Yablo 1992; Woodward 2008; Menzies 2008; List and Menzies 2009; cf. Zhong 2014). Such theories often deny the claim that physiological entities always cause bodily movements. This is their way out of the exclusion problem.

The second response is to give up nonreductive physicalism and deny the dependence claim. The result is that mental entities and physiological entities occupy the same level and one level does not depend on the other. Moreover, one way of rejecting dependence would be to say that mental entities and physiological entities are identical, rather than distinct, entities. So-called *mind–brain identity theories* comprise a family of views that exemplify this third response to the exclusion problem (Smart 1959; Armstrong 1968; Lewis 1970; Robb 2013). By giving up the dependence claim, these theories can hold on to the claims that both physiological and mental entities are causes. These theories could also accept the exclusion principle.[10]

The third response is to deny that mental entities are causes. When responding to the exclusion problem in this way, there are roughly two families of positions available. Both have in common the claim that the level of mental entities depends on the level of physiological entities. Furthermore, both families within this third option may accept the exclusion principle. But the two families differ in their view of the role of mental entities. Positions in the first family deny that mental entities have any causal or causal-like role to play at all. This is the position of *epiphenomenalism*, which denies causation by the mental and hence marks a significant departure from nonreductive physicalism. In contrast, positions in the second family accept that mental entities have a causal or causal-like role; they deny only that causation by mental entities is of the same kind as causation by physiological entities. Jackson and Pettit (1990, 115) distinguish between *causal efficacy* and *causal programing*. In their view, physiological entities are causally efficacious and mental entities causally program the effect. Both are causes, although in different ways. Similarly, Dretske (1991, 39–44) distinguishes between *triggering causes* and *structuring causes*.

[10] Although, strictly speaking, the principle does not apply to this position because the two causes – the mental entity and the physiological event – are one and the same.

What these theories have in common is that they distinguish between two kinds of cause. Although the theories thereby permit us to respond to the exclusion problem in a way that keeps much of nonreductive physicalism intact – they assume that mental entities are distinct from physiological entities and are determined by them – these theories weaken another central claim of nonreductive physicalism, namely that mental entities are causes (in the same sense in which physiological entities are causes).

The fourth response to the exclusion problem is to deny the exclusion principle itself. The resulting position looks appealing. It not only accepts the main tenets of nonreductive physicalism without distinguishing different kinds of causation, but the position furthermore contends that physiological entities are causes (in the same sense in which mental entities are causes) without facing a contradiction. Of course, a position in this fourth camp has to argue that there is something wrong with the exclusion principle. There are different ways of doing so.

Bennett (2003) argues that the exclusion principle is unjustified. Usually, the exclusion principle is justified using cases of systematic overdetermination. Bennett argues that this overlooks that the relevant causes in the exclusion problem operate on different levels. If the causes operate on different levels, then they *can* both be causes without being a case of systematic overdetermination. Hence, she denies the exclusion principle. Stoljar (2008; 2010, 217–20) argues that the exclusion principle is implausible. Once we clarify what is meant with "distinct," namely numerical distinctness (as opposed to modal or mereological distinctness), we find counterexamples to the principle (Stoljar 2008, 273). Finally, Kroedel (2015) shows that the exclusion principle is in fact inconsistent with widely held views about causation and the nature of mental entities in general.

4 The Disappearing Agent as an Exclusion Problem

We can now examine the analogy between the exclusion problem and the disappearing agent problem. In analogy to the responses to the exclusion problem, we can see underappreciated responses to the disappearing agent problem.

First of all, two assumptions need to be made explicit because they are required to formulate the disappearing agent problem as involving different levels. The first assumption is that individuals and mental entities are numerically distinct.[11] This assumption is compatible with the claim that mental entities are part of an individual because the claim is not that individuals and mental entities are mereologically distinct (cf. Stoljar 2008). So,

[11] I use "agents" and "individuals" interchangeably.

the individual may be a substance that bears intentions as properties or accidents. The second assumption is that individuals are associated with a different level than mental entities and that there is some ontological dependence relation between these levels. For example, understanding the dependence relation between levels as supervenience, you could say that the individual level supervenes on the intentional level.[12] In principle, the dependence relation could be reversed. Regardless of the direction of the dependence relation, the structure of the problem as an exclusion problem would be the same. Hence, I will simply speak of "different levels" in the formulation of the problem.

Are these assumptions plausible? I will return to this question when discussing the limits of the analogy. For now, note that proponents of the agent-causal approach seem to accept both assumptions. First, they assume that individuals are numerically distinct from their mental entities. Consider, for example, Taylor (1966, 111), who writes that an agent is "never identical with [what] ... is usually proposed as the 'real cause' of [the agent's] act, such as some intention or state of willing." Second, proponents of the agent-causal approach are likely to also accept the second assumption (Chisholm 1976, 199; Buckareff 2012, 106). They might reverse the dependence relation, given that they take the agent to be "ontologically irreducible" and more fundamental than events or mental entities. But since my formulation of the exclusion principle is neutral with respect to order of the ontological dependence, the disappearing agent problem as I construct it arises regardless of the order of ontological dependence.

With these assumptions in hand, the disappearing agent problem can be formulated in analogy to the exclusion problem for nonreductive physicalism as follows.

The Disappearing Agent Problem as an Exclusion Problem

Mental entities are causes. For each action, there are mental entities causing the action's result.

Dependence. Agents are at a different level than mental entities and are distinct from them.

Agents are causes. For each action, there is an agent causing the action's result.

Problem. These three claims are inconsistent with the exclusion principle.

Of course, there are differences between the exclusion problem and the disappearing agent problem formulated in this way. First, the relevant effects are different. For the exclusion problem, the effect is specifically behavior. For the disappearing agent problem, by

[12] It does not follow that whenever two individuals have the same mental entities they must be the same individual. Rather, the intentional level is associated with several other macroscopic events besides mental entities. The assumption could be that it is this level as a whole that supervenes on the intentional level.

contrast, the effects are actions or their results. Second, the relevant causes are different. The exclusion problem and the disappearing agent problem concentrate on causal claims on different pairs of levels. The exclusion problem in the philosophy of mind hardly ever considers agency on the individual level whereas the disappearing agent problem is precisely about the question of agency on the individual level. Third, the arguments are usually employed for different ends, which is reflected in that they tend to use the exclusion principle in different directions. In the philosophy of mind, the exclusion argument is deployed in an upwards direction by opponents of nonreductive physicalism who defend causation on the implementation level against causation on any higher level. In the philosophy of action, the disappearing agent problem is discussed by proponents of the agent-causal approach. They seek to defend causation on the individual level and either argue against causation on any lower level (e.g. Hornsby, Melden and Taylor, using the exclusion principle in a downward direction), or they recognize that their approach must defend itself against challenges from claims about causation on lower levels (e.g. Nagel and Velleman, using the exclusion principle in an upward direction).

Especially this last difference seems noteworthy. The fact that the exclusion principle tends to be used predominantly in an upward direction in the philosophy of mind but in a downward direction in the philosophy of action might indicate that two very different concerns are at stake in the respective debates.[13] This might, potentially, undermine my analogy claim and the hope that solutions to one problem can be transferred to the other. But the difference in how they are used amounts to a difference in the dialectic of the respective debates. Despite differences in the dialectic in which they are embedded, the arguments are structurally analogous. The analogy claim hence is not threatened by this difference. Nevertheless, the difference in concern might explain which avenues seemed promising and which solutions were salient or relevant in the respective debates. Because those discussing the disappearing agent problem in the philosophy of action tend to be proponents of an agent-causal approach, who consider this problem in defense of their account or who take this problem as a potential objection against their account, they have no immediate need to develop a different account of agent causation in response to the disappearing agent problem. But this is precisely what the analogy between the disappearing agent problem and the exclusion problem can provide.

More importantly, there are similarities between the exclusion problem and the disappearing agent problem. The two problems share a substantive assumption. Both appeal to the same exclusion principle. The two problems also have the same structure. Three assumptions lead to a contradiction with the exclusion principle. At best three of these four

[13] I am indebted to an anonymous reviewer for this observation.

claims can be accepted together; at least one has to be given up. Just as before there are, therefore, four ways of responding to the disappearing agent problem.

The first response to the disappearing agent problem is to deny that mental entities are causes. The resulting view is the well-known agent-causal approach claiming that only agents cause the results of their actions. Results of actions are not caused by mental entities or other things that are not under an agent's control (cf. Clarke 1993, 192). Although construing agent causation as substance causation is one way to spell out this view, other ways stand open as well. For example, it could be the *powers* of an agent that cause the result of an action (cf. Buckareff 2017).

Thanks to the analogy with the exclusion problem, proponents of such a view can now avail themselves of an additional argument. A so far underexplored avenue of research would attempt to extend proportionality theories of causation to support the agent-causal approach. The idea would be that the agent is the cause of an action because the agent is proportional to this effect. The relevant approach could operate with an explicit proportionality condition, following Yablo (1992), or the condition could be implicitly built into the approach as, for example, in the proposal of List and Menzies (2009). A crucial issue, as with existing accounts of agent-causation, is to explain what agents are. What is the ontological kind of agents? And how do proportionality considerations apply to this ontological kind? At any rate, this avenue explores the view that the level at which the cause of an action is located is the individual level.

The second response is to deny that agents and their mental entities occupy different levels. This denies the dependence claim. At first, it might seem hard to make out the resulting view. But perhaps it is best to understand this option by analogy with the corresponding view in philosophy of mind. By analogy with mind–brain identity theories, we get agent–intention identity theories. Although a theoretical option, this view is unlikely to strike many as plausible. The view will be especially unattractive to proponents of the agent-causal approach because such an identity theory would abandon agent-causation in reducing it to event-causation. But there are more attractive alternatives besides crude agent-intention identity theories. The idea would be that an agent is not identical to *all* her intentions or mental entities, but rather to a certain privileged set of intentions that are involved in causing the results of an action. It is in virtue of the contribution of these privileged intentions that an agent counts as active.

There exist several such *identification theories of agency*, although they differ in how they identify the set of privileged intentions. The view that the agent is identical to a certain set of mental entities is put forward by Velleman (1992, 474–75). Velleman aims at a "reduction of agent-causation," writing that "we should look for events and states to play the

role of the agent." In particular, he focuses on "mental events and states that are functionally identical to the agent, in the sense that they play the causal role that ordinary parlance attributes to him." So Velleman identifies the agent with those mental entities that play a certain causal and functional role. Another identification theory of agency is put forward by Frankfurt (1971). Frankfurt focuses on those intentions that an agent wants to have, instead of those that play a certain functional role. He gives an "analysis of the concept of a person" in terms of those intentions. Watson (1975, 215–16) can be seen as defending a similar identification theory of agency. He argues that an agent is identical with certain values or her overall "valuational system." He writes that "one cannot dissociate oneself from all normative judgments without forfeiting all standpoints and therewith one's identity as an agent." In other words, the privileged set of mental entities with which Watson identifies the agent is certain values or normative judgments. Finally, more to add to rather than to complete this list, consider Shoemaker (2015), who offers another reductive analysis of, if not the agent, then at least the agent's character in terms of commitments and care tendencies. Although I omit many details of these accounts here, I find it helpful to see what unites such very different accounts as potential responses to the disappearing agent problem. Each of these different accounts can be seen as responding to the disappearing agent problem by denying the dependence assumption.

The third response to the disappearing agent problem is to deny that agents are causes. This is how some proponents of the event-causal approach respond to the disappearing agent problem. This line of response has the appeal of lining up a standard view of the metaphysics of agency with a standard view of the metaphysics of causation. But, of course, this is not a line of response that proponents of agent-causation could take, because the resulting view jeopardizes the main tenet of that approach. Also, it is not clear that aligning with a standard is an important theoretical virtue. At any rate, in the context of the exclusion problem in philosophy of mind, the position that corresponds to this response would be *epiphenomenalism*. And as in philosophy of mind so in philosophy of action, this position seems to deny something obvious.

Besides epiphenomenalism, there is a second, crucially different family of positions that also corresponds to the third way of responding to the argument. Positions in this second family contend that both agents and events are causes, but they deny that they are causes of the same kind. Instead, agents and events play different causal roles. The distinction between *causal programing* and *causal efficacy* from Jackson and Pettit (1990) or the distinction between *structuring causes* and *triggering causes* from Dretske (1991) could be applied to the disappearing agent problem. Clarke (1993, 194) seems to put forward a reconciliation between agent-causation and event-causation in this spirit. He writes:

> [W]hat an agent directly causes … is her acting on (or for) certain of her reasons rather than on others, and her acting for reasons ordered in a particular way …. Her acting for that ordering of reasons is itself a complex event, one that consists, in part, of her behavior's being caused by those reasons.

The causal role of the agent here — that she causes her acting "for reasons ordered in a particular way" — seems to be of a structuring kind. The acting itself, in turn, can be seen as a triggering cause, an "event" of those reasons causing the result of the action.

Similarly, Steward (2016) accounts for agent-causation by seeing the agent as occupying a special causal role, namely, that of sustaining the process of action.[14] A related view in the vicinity could be developed from the insight that we can distinguish two forms of causal efficacy for substances and their powers (Whittle 2016; Buckareff 2017). Each of these views slightly weakens a central tenet of the agent-causal approach – that both agents and events cause actions in the same way – but retains its spirit.

The fourth response to the disappearing agent problem is to *deny the exclusion principle*. This response is straightforward because it is parallel to the response in the context of the exclusion problem. Each time, the same exclusion principle is at stake. The objections against the exclusion principle by Bennett (2003) or Stoljar (2008; 2010) apply directly. In fact, some proponents of substance causation are ecumenical and allow for both substances and events to be causes (Whittle 2016, 2). The resulting position looks attractive because on this position, an approach may be both agent-causal and event-causal at the same time. This illustrates how arguments that have been developed elsewhere may be applied to the disappearing agent problem.

But there are limits to this analogy. First, there is a general question about the ontological kinds involved in formulating the disappearing agent problem as an exclusion problem. The foregoing formulation said nothing specific about what individuals and intentions are. But an important issue concerns what kind of thing we find on each of the different levels. The literature varyingly uses "states," "events," "processes," and "properties" to describe intentions and entities on the intentional level and the implementation level. Specifying ontological categories is a demand of ontological seriousness (Buckareff 2012, 106). I have framed the discussion throughout this paper in terms of "entities" or "events," assuming that the differences do not matter for the purposes here.

Second, there is more specifically the question of what individuals are. How do individuals fit into this picture of levels? Unless the picture of levels that I provide here says

[14] This answer, of course, goes hand-in-hand with a certain metaphysics of actions, namely, seeing them as processes (Steward 2012a).

more about the ontological kind of individual agents, it will be difficult to fully evaluate the assumption that the individual level depends on the intentional level (or the other way around). Although this is a limitation, this issue can largely be set aside for present purposes. First of all, the ontology of agents as well as the notion of dependence between levels have been investigated at length elsewhere. More importantly, my aim here has only been to make a point about dialectical structure. I argued for the fruitfulness of an alternative picture that sees the disappearing agent problem and the exclusion problem as analogous.

5 Agent Causation as Proportional Causation

Identifying an analogy between two problems on which the literature is largely disjointed is progress by theoretical unification. In addition, I suggested that this analogy can spur theoretical innovation. I discussed a taxonomy of potential solutions to make plausible in a systematic way how gaps can be identified, and solutions might be transferred. But this was all rather abstract, and it left open many ontological questions. Now I should demonstrate more concretely how the analogy helps to transfer solutions from one problem to the other. To do so, I work out in more detail how proportionality theories of causation can help make a case for agent causation. This theoretical option tries to make good on two claims: that agents cause results of actions, and that mental entities do *not* cause results of actions. Specifically, I explore how the theory of List and Menzies (2009) may support such an agent-causal approach.

Let me start by describing first how a proportionality theory of causation addresses the exclusion problem in philosophy of mind (for details see List and Menzies 2009). Let causation, for now, be understood as a relation between binary variables encoding such things as whether a property is instantiated or whether an event occurs. When a person reaches for a cup, is this behaviour B caused by the presence of a mental entity M or a neurophysiological implementation N? A proportionality theory of causation argues that, in such exclusion problem situations, a causal relation obtains between M and B but *not* between N and B.

To achieve this result, List and Menzies (2009) make two important theoretical proposals. First, they propose to understand causation as difference-making with the following truth conditions.

> The presence of F makes a difference to the presence of G in the actual situation if and only if (i) if any relevantly similar possible situation instantiates F, it instantiates

G; and (ii) if any relevantly similar possible situation instantiates $\neg F$, it instantiates $\neg G$ (List and Menzies 2009, 482).

List and Menzies (2009) evaluate these truth conditions in Lewis' (1973) semantics for counterfactuals based on a distance relation between possible worlds that may change with context. In the truth conditions above, they therefore replace "situation" with "world" and understand the two conditions as counterfactuals. List and Menzies moreover amend the semantics by requiring that the system of spheres of possible worlds encoded by the distance relation is *weakly centred*. That is, the smallest sphere of possible worlds around the actual world *w* may contain more than one world. Otherwise, condition (i) of the truth conditions of difference-making would be trivially true. This is because the actual world is an *F*-world and a *G*-world (that is, *F* are *G* are present in the actual world) and the alternative *strong centring* assumption would imply that the only *F*-world closest to *w* is *w* itself (but since this world, the actual world *w*, is also a *G*-world, condition (i) is true, according to Lewis' truth conditions for counterfactuals).

Second, List and Menzies (2009) characterize the property of causal relations being *realization sensitive* ("realization" here is one form of ontological dependence discussed earlier). Specifically, where the presence of a mental entity *M* is realized by the presence of a neural state *N*, the causal relation between *M* and *B* is realization-sensitive if, "in all those *M*-worlds that are closest $\neg N$ worlds (such that *M* has a different realizer), *B* is no longer present" (2009, 493). Intuitively, the presence on *B* is sensitive to the presence of the realizer *N*, just as much as it is sensitive to the presence of *M* itself. Interestingly, List and Menzies (2009) show that if a higher-level causal relation between *M* and *B* is instead realization-*in*sensitive, that is, if the relation continues to hold even under small changes in how *M* is realized, then this causal relation rules out lower-level causes of *B* consistent with the exclusion principle. In a slogan, proportional causal relations are those that are realization insensitive. As List and Menzies (2009) argue, this result is obtained from the truth conditions of difference-making operationalized in the amended semantics of Lewis (1973). Similar observations have been offered by Lewis (1987) and Woodward (2006).

Applied to the exclusion problem, the following picture emerges. Mental entities *M* can be realized by different neural states N_1, N_2, N_3, etc., of which N_1 actually realizes *M*. The causal relation between *M* and *B* is realization insensitive. Specifically, it meets condition (i) of the truth conditions of difference-making: In all nearby worlds in which *M* is present, *B* is present. The same is *not* the case for N_1. Even where N_1 is not present, and *M* is instead realized by N_2, *B* is still present. Therefore, *M* is a cause of *B*, but N_1, is not.

This idea of causation as difference-making allows us to construct a novel agent-causal approach to agency. The proposal of List and Menzies (2009) transfers over to the

disappearing agent problem as follows. The idea would be that just as the causal relation between mental entity *M* and behaviour *B* is realization-insensitive, so is the causal relation between the presence of some agent *A* and some result of action *R*. A robust causal relation obtains between *A* and *R* but not between *M* and *R*. With the truth conditions of List and Menzies (2009), this is the case when the distance relation over possible worlds is such that in all closest *A*-worlds around *w*, *R* is present (some of these *A*-worlds are worlds where *M* is present, some are worlds where *M* is absent) and in all closest worlds around *w* in which *A* is not present, *R* is also not present. In result, *A* causes the result of action *R* but the mental entity *M* that realizes the agent does not cause *R*.

This account, so far, identifies only what the distance relation *must* be like for agent causation as a form of proportional causation. But the question is whether the distance relation of possible worlds is *actually* such that the causal relation between *A* and *R* is realization-insensitive while the relation between *M* and *R* is realization-sensitive. Whether this is the case depends on the ontological assumptions of the account, especially on the ontology of individuals. First, what does the variable *A* actually stand for? Second, how does it make sense to say that a mental entity *M* realizes *A*? Third, is *M* still causing something? After all, when it came to the exclusion problem, the proportionality theory of causation implied that *M* is a cause (but *N* is not). If the same theory now implied that *M* is not a cause, the theory gives with one hand and takes with another.

I will try to answer these three ontological issues in turn: what the causal variable *A* stands for, how *M* can be said to realize *A*, and what *M* still causes. Given the depth of the issues and the limitations of space here, my answers will be necessarily suggestive and speculative, but they should suffice to illustrate more concretely the potential opportunities of exploiting the analogy between the exclusion and the disappearing agent problem.

First, variable *A* stands for the agent at the time of her acting. But saying that *A* stands for the "agent at a time" by itself is, of course, not very illuminating and raises long-standing issues of personal identity and, perhaps more specifically, of individual ontology. Several options are available as referents of "agent at the time of acting." Among other things, human agents could be animals, brains, property-bundles or temporal parts or stages (Olson 2007). At any rate, whatever we take agents to be, they should be substances in order to keep the present proposal in line with the tradition of agent-causal approach. But again, different theoretical options present themselves to understand the ontological category of substance, including understanding substance as property bundles (Robinson 2020). On the view that I favour, agents are property bundles — and substances, as an ontological category, are understood as property bundles as well. As a result, agents are numerically distinct but consist, in part, of mental entities. Agents change over time, depending on what mental

entities they instantiate as their parts.[15] An agent at a time is a property bundle at a time and, hence, a substance.

Second, the mental entity *M* realizes an agent at a time *A* only in the sense that *M* is a part of *A* (and that all entities on the intentional level, which *M* occupies, determine the entities on the individual level, which *A* occupies). That *M* occurs is an event, for example, that some agent intends to have coffee. This ontology of mental entities is, however, the less significant part of this account because, as an agent-causal approach, the account needs to afford little concern for mental entities. Nevertheless, just to complete the ontological picture, I assume that *M* is an event, insofar as events are considered the least suspicious category of causal relata and often seen as such in the literature on mental causation.

Third, the mental event *M* is still a cause of some other event. Of course, *M* does not cause the results of action *R* (which are caused by *A*). Proportionality theories of causation generally seek to be consistent with the exclusion principle as stated above. Therefore, because this account contends that *A* causes *R*, by the exclusion principle, *R* cannot also be caused by *M*. Instead, a proportionality theory of agent causation will suggest different effects on different levels. Whereas *A* causes *R* and both occupy the individual level, *M* causes some other event *B* and both occupy the intentional level. Interpretationally, *B* could be some macro behaviour or conscious movement that realizes but is distinct from the result of the action *R*. Likewise, one level further down *N*, an event on the implementation level, may cause a further event *Y* on the implementation level. Just as *A* is realized by *M*, which is realized by *N* in turn, so *R* may be realized by *B*, which may be realized by *Y* in turn. Even without an ontology of levels, the literature on philosophy of action contains extant accounts of such compositional theories of action individuation on which "larger" To my knowledge, no such agent-causal approach to agency, as the one I just sketched, based on the idea of proportional causation has been defended in full. But several theories of proportional causation are available. Like the account just sketched, these theories may require further assumptions about the ontology of agents to be plausible. But several theories of personal identity and human agency ontology are available as well. I hope this sketch can hence serve as an example of how the analogy between the exclusion problem and the disappearing agent problem can be fruitful and can be put to use.

[15] To account for action at a time, I do not need to assume that agents are identical over time and I therefore do not sketch this part of the account here.

6 Conclusion

This paper is about a hypothesis that had been entertained before but whose correctness has not yet been put to a test. Is the disappearing agent problem an exclusion problem? I have first described each problem and then developed a dialectic on which the two problems are analogous. Moreover, the disappearing agent problem and the exclusion problem share a central assumption, the exclusion principle. I have rendered the two problems with the general philosophical picture of ontological levels (cf. List 2019). On this picture, the exclusion problem is, roughly, the disappearing agent problem just a few levels down.

The contribution of this paper is not only to work out this analogy and the substantive overlap between the disappearing agent problem and the exclusion problem but also to illustrate that these insights can be fruitful. Once we see the analogous structure of the two problems, we can see a taxonomy of existing solutions and we are given hints of where to look for new responses. Specifically, the response of denying the exclusion principle has received comparatively little attention in the disappearing agent problem literature. Furthermore, proportionality theories of causation could be a novel form of agent-causation. Finally, a distinction between different kinds of causes might form the basis of a position that reconciles agent-causation and event-causation without having to reject the exclusion principle. Overall, the dialectic that I offer yields not only a taxonomy for existing arguments, it allows us to see, in a systematic way, several ways of reconciling agent-causation and event-causation and accepting both.

Acknowledgements

Ryan Cox and Daniel Stoljar helped clarify my thinking during a visit to the Research School of Social Sciences at the Australian National University (ANU) in 2013. The paper also benefitted from comments and discussions with Richard Bradley, Richard Holton, Sebastian Köhler, Christian List, Wlodek Rabinowicz, Jesse Saloom, and Ying Shi and discussions in Geert Keil's colloquium at HU Berlin in 2016. Finally, I was able to improve the paper's substance or exposition in response to comments from anonymous reviewers at this and at least three other journals.

References

Aguilar, Jesús Humberto, and Andrei A. Buckareff, eds. 2010. *Causing Human Actions: New Perspectives on the Causal Theory of Action*. Cambridge, Mass.: The MIT Press.

Alvarez, Maria, and John Hyman. 1998. “Agents and Their Actions.” *Philosophy* 73 (2): 219–245.

Armstrong, D. M. 1968. *A Materialist Theory of the Mind*. London: Routledge & Kegan Paul.

Bennett, Karen. 2003. “Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It.” *Noûs* 37 (3): 471–97. https://doi.org/10.1111/1468-0068.00447.

———. 2007. “Mental Causation.” *Philosophy Compass* 2 (2): 316–37. https://doi.org/10.1111/j.1747-9991.2007.00063.x.

———. 2008. “Exclusion Again.” In *Being Reduced: New Essays on Reduction, Explanation, and Causation*, edited by Jakob Hohwy and Jesper Kallestrup, 280–305. Oxford: Oxford University Press.

Bishop, John. 2007. “On the Importance of Reconciling Two Models of Causality in Agency.” In *Action in Context*, edited by Anton Leist, 154–62. Berlin: Walter de Gruyter.

Brent, Michael. 2017. “Agent Causation as a Solution to the Problem of Action.” *Canadian Journal of Philosophy* 47 (5): 656–73. https://doi.org/10.1080/00455091.2017.1285643.

Buckareff, Andrei A. 2012. “How Does Agent-Causal Power Work?” *The Modern Schoolman* 88 (1/2): 105–121.

———. 2017. “A Critique of Substance Causation.” *Philosophia* 45 (3): 1019–26. https://doi.org/10.1007/s11406-017-9855-7.

———. 2018. “I’m Just Sitting around Doing Nothing: On Exercising Intentional Agency in Omitting to Act.” *Synthese* 195 (10): 4617–35. https://doi.org/10.1007/s11229-017-1424-x.

Chisholm, Roderick M. 1976. “The Agent as Cause.” In *Action Theory*, edited by Myles Brand and Douglas Walton, 199–211. Synthese Library 97. Springer Netherlands. http://link.springer.com/chapter/10.1007/978-94-010-9074-2_12.

Clarke, Randolph. 1993. “Toward A Credible Agent-Causal Account of Free Will.” *Noûs* 27 (2): 191–203. https://doi.org/10.2307/2215755.

———. 2017. “Free Will, Agent Causation, and ‘Disappearing Agents.’” *Noûs*, May. https://doi.org/10.1111/nous.12206.

Dretske, Fred. 1991. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, Mass.: The MIT Press.

Frankfurt, Harry G. 1971. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68 (1): 5–20. https://doi.org/10.2307/2024717.

Goldman, Alvin I. 1970. *A Theory of Human Action*. Princeton, N.J.: Princeton University Press.

Gutting, Gary. 2009. *What Philosophers Know: Case Studies in Recent Analytic Philosophy*. Cambridge: Cambridge University Press.

Hornsby, Jennifer. 2004. "Agency and Actions." In *Agency and Action*, edited by John Hyman and Helen Steward. Royal Institute of Philosophy Supplement 55. Cambridge: Cambridge University Press.

Jackson, Frank, and Philip Pettit. 1990. "Program Explanation: A General Perspective." *Analysis* 50 (2): 107–17. https://doi.org/10.2307/3328853.

Kim, Jaegwon. 1989. "The Myth of Nonreductive Materialism." *Proceedings and Addresses of the American Philosophical Association* 63 (3): 31–47. https://doi.org/10.2307/3130081.

———. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, Mass.: The MIT Press.

———. 2005. *Physicalism, Or Something Near Enough*. Princeton, N.J.: Princeton University Press.

Kroedel, Thomas. 2015. "Dualist Mental Causation and the Exclusion Problem." *Noûs* 49 (2): 357–375. https://doi.org/10.1111/nous.12028.

Lewis, David. 1970. "An Argument for the Identity Theory." *Journal of Philosophy* 63 (2): 17–25.

———. 1973. *Counterfactuals*. Oxford: Wiley-Blackwell.

———. 1987. "Postscripts to 'Causation.'" In *Philosophical Papers: Volume II*, 172–213. Oxford: Oxford University Press.

List, Christian. 2019. "Levels: Descriptive, Explanatory, and Ontological." *Noûs* 53 (4): 852–83. https://doi.org/10.1111/nous.12241.

List, Christian, and Peter Menzies. 2009. "Non-Reductive Physicalism and the Limits of the Exclusion Principle." *The Journal of Philosophy* 106 (9): 475–502.

Lowe, E. Jonathan. 2008. *Personal Agency: The Metaphysics of Mind and Action*. Oxford: Oxford University Press.

Malcolm, Norman. 1968. “The Conceivability of Mechanism.” *The Philosophical Review* 77 (1): 45–72. https://doi.org/10.2307/2183182.

Mayr, Erasmus. 2011. *Understanding Human Agency*. Oxford: Oxford University Press.

Melden, A. I. 1961. *Free Action*. New York: Humanities Press.

Menzies, Peter. 2008. “The Exclusion Problem, the Determination Relation, and Contrastive Causation.” In *Being Reduced: New Essays on Reduction, Explanation, and Causation*, edited by Jakob Hohwy and Jesper Kallestrup, 196–217. Oxford: Oxford University Press.

Mumford, Stephen, and Rani Lill Anjum. 2011. *Getting Causes from Powers*. Oxford: Oxford University Press.

Nagel, Thomas. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

O’Connor, Timothy. 2005. “Freedom with a Human Face.” *Midwest Studies In Philosophy* 29 (1): 207–27. https://doi.org/10.1111/j.1475-4975.2005.00113.x.

Olson, Eric T. 2007. *What Are We?: A Study in Personal Ontology*. Oxford University Press.

Pereboom, Derk. 2014a. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

———. 2014b. “The Disappearing Agent Objection to Event-Causal Libertarianism.” *Philosophical Studies* 169 (1): 59–69. https://doi.org/10.1007/s11098-012-9899-2.

Robb, David. 2013. “The Identity Theory as a Solution to the Exclusion Problem.” In *Mental Causation and Ontology*, edited by Sophie Gibb, E.J. Lowe, and R.D. Ingthorsson, 215–32. Oxford: Oxford University Press.

Robinson, Howard. 2020. “Substance.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2020. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2020/entries/substance/.

Schlosser, Markus. 2015. “Agency.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2015. http://plato.stanford.edu/archives/fall2015/entries/agency/.

Shoemaker, David. 2015. *Responsibility from the Margins*. Oxford: Oxford University Press.

Smart, J. J. C. 1959. “Sensations and Brain Processes.” *The Philosophical Review* 68 (2): 141–56. https://doi.org/10.2307/2182164.

Steward, Helen. 2012a. “Actions as Processes.” *Philosophical Perspectives* 26 (1): 373–388. https://doi.org/10.1111/phpe.12008.

———. 2012b. *A Metaphysics for Freedom*. Oxford: Oxford University Press.

———. 2016. “Making the Agent Reappear: How Processes Might Help.” In *Time and the Philosophy of Action*, edited by Roman Altshuler and Michael J. Sigrist, 289. Routledge Studies in Contemporary Philosophy. London: Routledge.

Stoljar, Daniel. 2008. “Distinctions in Distinction.” In *Being Reduced: New Essays on Reduction, Explanation, and Causation*, edited by Jakob Hohwy and Jesper Kallestrup. Oxford: Oxford University Press.

———. 2010. *Physicalism*. Oxford: Routledge.

Stout, Rowland. 1996. *Things That Happen Because They Should: A Teleological Approach To Action*. Oxford: Oxford University Press.

———. 2005. *Action*. Chesham: Acumen.

———. 2007. “Two Ways to Understand Causality in Agency.” In *Action in Context*, edited by Anton Leist, 137–53. Berlin: Walter de Gruyter.

———. 2012. “Mechanisms That Respond to Reasons: An Aristotelian Approach to Agency.” In *Human Destinies: Philosophical Essays in Memory of Gerald Hanratty*, edited by Fran O’Rouke. Notre Dame: Notre Dame University Press.

Taylor, Richard. 1966. *Action and Purpose*. Englewood Cliffs, N.J.: Prentice-Hall.

Velleman, J. David. 1992. “What Happens When Someone Acts?” *Mind* 101 (403): 461–81. https://doi.org/10.2307/2253898.

———. 2020. *On Being Me: A Personal Invitation to Philosophy*. Princeton University Press.

Watson, Gary. 1975. “Free Agency.” *The Journal of Philosophy* 72 (8): 205–20. https://doi.org/10.2307/2024703.

Whittle, Ann. 2016. “A Defense of Substance Causation.” *Journal of the American Philosophical Association* 2 (1): 1–20. https://doi.org/10.1017/apa.2016.1.

Woodward, James. 2006. “Sensitive and Insensitive Causation.” *The Philosophical Review* 115 (1): 1–50.

———. 2008. “Mental Causation and Neural Mechanisms.” In *Being Reduced: New Essays on Reduction, Explanation, and Causation*, edited by Jakob Hohwy and Jesper Kallestrup, 218–62. Oxford: Oxford University Press.

Yablo, Stephen. 1992. “Mental Causation.” *The Philosophical Review* 101 (2): 245–80. https://doi.org/10.2307/2185535.

Zhong, Lei. 2014. “Sophisticated Exclusion and Sophisticated Causation.” *Journal of Philosophy* 111 (7): 341–60. https://doi.org/10.5840/jphil2014111724.