Canadian Journal of Philosophy, 2015
Vol. 45, No. 3, 300–325, http://dx.doi.org/10.1080/00455091.2015.1080045

Routledge
Taylor & Francis Group

Actions and accidents

David Horst*

Martin Buber Society of Fellows, Mandel School for Advanced Studies in the Humanities, Hebrew University of Jerusalem, Room 321/322, Mt. Scopus, Jerusalem 91905, Israel

(*Received 13 November 2014; accepted 3 August 2015*)

In acting intentionally, it is no accident that one is doing what one intends to do. In this paper, I ask how to account for this non-accidentality requirement on intentional action. I argue that, for systematic reasons, the currently prevailing view of intentional action – the Causal Theory of Action – is ill-equipped to account for it. I end by proposing an alternative account, according to which an intention is a special kind of cause, one to which it is essential that it represents its effect.

Keywords: intentional action; causal theory of action; reasons as causes; Donald Davidson; G.E.M. Anscombe

1. Introduction

In 'Actions, Reasons, and Causes' Donald Davidson famously claims that reasons are causes (See Davidson 1963). The argument for this claim is as simple as it is convincing: in order to act intentionally it is not enough to move in *accordance* with one's reason, rather one's movement must be *caused* by the reason with which it accords. The causal condition is required to avoid collapsing the distinction between intentional action, on the one hand, and, on the other, movement that just happens to be in accord with one's reason by mere *accident*.

Davidson, as everyone knows, took the reason for which one performs an intentional action to be a belief-desire pair. But the core of Davidson's argument does not depend on any specific view about reasons for acting. It does not even depend on the view that acting intentionally is acting for a reason. To bring out the core of the argument, we may replace 'reason' with 'intention' and stay non-committal as to what is involved in intending to do something beyond what is contained in Davidson's argument itself. (The corresponding slogan would thus be: 'intentions are causes'.)[1] Davidson's argument can, then, be taken to contain the following two ideas. The first is what we may call the *non-accidentality requirement* on intentional action:

*Email: david.horst@mail.huji.ac.il

(Non-Accidentality) If one is acting intentionally, it is *no accident* that one is doing what one intends do.

The second idea is a specification of the role of one's intention in acting intentionally:

(Intention) In acting intentionally, one's intention is the *cause* of the movement it *represents*.[2]

The main move in the argument, then, is to claim that we can account for (Non-Accidentality) in terms of (Intention): in acting intentionally, it is no accident that one is doing what one intends to do if and because in so acting one's intention is the cause of what it represents.

I think this is exactly right. Intentional action is subject to a non-accidentality requirement; and, to meet this requirement, we should conceive of an intention as the cause of what it represents. However, I shall argue that, in light of certain widespread assumptions in contemporary action theory, it is not at all clear how we *can* hold on to this insight. This comes out when we consider the standard way of understanding (Intention) – the so-called causal theory of action (CTA), which attempts to understand this idea in terms of a reductive account of intentional action (I will sketch the CTA and its underlying assumptions in Section 2). For, as I will argue, if we conceive of an intention's causal power along the lines of the CTA, the intention's causality will be at best necessary, but *not* sufficient for (Non-Accidentality) (Section 3). But, as I shall further argue, there is good reason to think that, in the end, a satisfactory account of (Non-Accidentality) depends on our ability to understand an intention's causal power in such a way that it does in fact provide a *sufficient* account for (Non-Accidentality). As we will see, this claim turns on the argument that supplementing the CTA's causal condition with certain *non*-causal – especially, modal – conditions won't help with securing the relevant sort of non-accidentality (Sections 4 and 5).

The upshot of this argument may be surprising, since Davidson's argument is usually taken to provide the strongest support for adopting the CTA. Yet, it is the aim of this paper to undermine this widespread assumption by arguing, first, that the CTA is neither identical with nor entailed by (Intention), and, more importantly, that for systematic reasons the CTA is ill-equipped to account for (Non-Accidentality).

The shortcomings of the CTA will then motivate the exploration of an alternative account of (Non-Accidentality). More specifically, I shall suggest that a satisfactory account of this sort requires us to conceive of an intention as a *special kind* of cause, one to which it is *essential* that the intention is a representation of its effect (Section 6). As we will see, this view is importantly different from the CTA's conception of an intention's causality. But, if I am right, this is what is needed to understand an intention's causal power in such a way that it is sufficient to account for (Non-Accidentality). However, my

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

aim in this paper is not to present a fully fledged alternative to the CTA, rather it is to argue for the need of developing such an alternative on the ground that this is what is needed to account for the non-accidentality requirement.

Before I begin, a quick word on terminology. As I use the term, an intention is the 'cause' of what one is doing in the sense that it provides an explanation of the *existence* or *actuality* of one's movement. I take it that this sense is central to Davidson's argument. For, on the face of it, his idea seems to be that, in order to exclude the possibility that one's movement just happens to be in accord with one's intentions by mere accident, the intention has to be the cause of one's movement precisely in the sense that it explains the existence or actuality of one's movement. When I speak of an intention as the 'representation' of what one is doing, roughly and minimally I mean that, in intending to do A, one deploys the concept 'doing A', under which one's movement falls (or with which it accords, or to which it conforms, or which it realizes). When I speak of the content of one's intention, I mean the intention insofar as it deploys or contains the relevant concept.

2. The causal theory of action

2.1. The CTA's way of accounting for (Intention)

Let me begin with a brief discussion of the CTA and how it understands (Intention), i.e. the idea that, in acting intentionally, one's intention is the cause of what it represents. As I understand it, the CTA purports to give a *reductive* account of intentional action. This is an attempt to understand intentional action by stating necessary and jointly sufficient conditions for the existence of such action; it is reductive in the sense that the notions deployed in these conditions are taken to be more fundamental than and independently intelligible of intentional action itself. So, if we apply this sort of explanation to (Intention), we arrive at the following account of intentional action:

> (CTA) An agent's movement is an intentional action if (and because): (i) her intention is a *representation* of her movement and (ii) her intention is the *cause* of her movement.

This is the core of the sort of account that I call 'the CTA'.[3] To be sure, how both conditions are to be spelled out in any more detail is highly controversial. For instance, proponents of the CTA notoriously differ over what kind of mental state(s) must be part of a successful account of intentional action (e.g. whether or not intention can be reduced to some combination of belief and desire). They also differ as to which further condition(s) may be necessary to yield a sufficient account of intentional action. (I will discuss some proposals below.) But however significant the differences between particular versions of the CTA may be, they are unified by being precisely *versions* of the basic idea that, in acting intentionally, the agent's intention must both represent and cause

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

her movement, where the relevant notions of causality and representation are taken to be more fundamental than and independent of intentional action. What's important for my argument in this paper is, if you like, the general *structure* of the account that the different versions of the CTA have in common, not the differences in *content* by which they are distinguished. So let me say a little more about the structure of this account.

First of all, the above statement of the CTA looks so close to (Intention) that one may wonder what the alleged difference between them really comes to. To see the difference, it is important to appreciate that the CTA applies a reductive account to (Intention), but that, just by itself, (Intention) doesn't force such an account on us. To illustrate this, consider a more familiar point from epistemology: one may acknowledge that knowledge involves justification, truth, and belief, without *thereby* being committed to a reductive account of knowledge, where this means that justification, truth, and belief are more fundamental than the notion of knowledge itself. (An example is Williamson's [2000] knowledge-first approach.) Similarly, one may acknowledge that intentional action involves causation (as (Intention) claims), without *thereby* being committed to the view that causation is the more fundamental notion here (as the CTA claims). I will say more about what motivates the CTA's reductive approach below; for now it's just important to note the difference between (Intention) and the CTA.

Furthermore, it is crucial to my argument that the CTA takes the notions of causality and representation it deploys in its account to be *intrinsically* independent of one another. By 'intrinsically independent' I mean that, just as such, the obtaining of one of these relations – the causal and the representational relation – doesn't imply or require the obtaining of the other. That is to say, just by itself, the fact that a movement is caused by the intention doesn't imply or require that the representational condition is satisfied as well. It may be that the movement caused by the intention is not the movement represented by the intention. And, conversely, it may be that the intention represents a movement without being its cause. After all, if in fact causality implied or required representation (or *vice versa*), both conditions wouldn't be *separately* satisfiable. Consequently, causality and representation couldn't obtain – or could obtain only in some necessarily defective form – outside their unity in intentional action. But then, it seems these notions would no longer be more fundamental than what they purport to explain, and thus, the account would no longer be reductive in the relevant sense. Hence, I take it that the CTA's reductive ambition commits its proponents to the intrinsic independence of causality and representation.

To be sure, this is not to say that the CTA must deny just *any* connection between causality and representation. On the contrary, as I shall discuss at great length in Sections 4 and 5, proponents of the CTA often require that, for there to be an intentional action, there must also be a *modal* connection

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

between an intention's causality and its content, such that if the intention's content had been different, it would have caused a different movement. What I take the CTA to be committed to is just that the connection cannot be intrinsic in the sense explained. For, if it was, causality and representation would be of a *special kind*, the kind that had to be understood through their unity in intentional action. And this, I take it, is incompatible with the reductive aim of explaining intentional action in terms more fundamental than what they purport to explain.

Davidson, as we have noted, identified the mental cause of an intentional action with a belief-desire pair. On his view, we thus get the following version of the CTA:

> The action on the one hand, and the belief-desire pair which give the reason on the other, must be related in two very different ways to yield an explanation. First, there must be a logical relation. Beliefs and desires have a content, and these contents must be such as to imply that there is something valuable or desireable about the action […]. Second, the reasons an agent has for acting must, if they are to explain the action, be the reasons on which he acted; the reasons must have played a *causal* role in the occurrence of the action. (Davidson 1982, 173)

The logical and the causal relation of which Davidson speaks correspond to the two core conditions of the CTA, the representational and the causal condition. That both relations are 'very different', as Davidson says, reflects the view that they are intrinsically independent. Furthermore, Davidson states that the representational relation has a *normative* significance: the agent's thoughts represent her movement as something desirable or valuable to do.[4] On Davidson's view, then, for a movement to be an intentional action it must stand in two 'very different' relations to the agent's thought: a representational (or rationalizing, or justificatory) and a causal relation.

2.2. *Assumptions of the CTA*

It is sometimes suggested or just assumed that (Intention) implies the truth of some version of the CTA: maybe not Davidson's belief-desire-model, but then some other more sophisticated version of the CTA. However, as I have already noted, I think the CTA is neither identical with nor entailed by (Intention), i.e. the bare idea that, in acting intentionally, one's intention is the cause of what it represents. We have to distinguish between (Intention) and the particular way in which the CTA attempts to account for it. However, one may wonder why the transition from the one to the other seems so natural and even inevitable to so many people. This becomes clearer if we consider what I take to be the two main motivating assumptions of the CTA. The first one concerns the notion of *causality*, the second concerns the notion of *content*. I will briefly discuss each one in turn.

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

The first assumption is sometimes explicitly stated by proponents of the CTA: 'The [CTA's] perspective on agency is largely motivated by the worthy desire for a *uniform* understanding of causality' (Bishop 1983, 63; my italics).[5] On a *uniform* understanding of causality, there is only one kind of causality. The causality that binds my action to my intention is of the same kind as the causality that governs the movements of non-human animals or even inanimate objects. When I explain my action by saying 'I am doing A *because* …' and cite my intention or my reason for doing A, the 'because' refers to a causal connection which as such is no different from the one I refer to in explaining: 'The house is burning down *because* it was hit by lightening'. Thus, on this view, the causal role of my intention is specifiable using a notion of causality that also binds fire to lightening, or, notoriously, the movements of one billiard ball to the movements of another. Obviously, then, it can make no intrinsic difference to the causal relation in which my intention stands to what I am doing that the items thus connected also stand in a relation of representation (or rationalization). A prominent example of a uniform understanding of causality is Davidson's (1970) nomological view of causality, according to which it is part of the very idea of causality that causally connected events can be subsumed under a natural law (under some description). Thus, ultimately, any causal explanation is made true by the same fundamental kind of causality, the one described in the relevant natural law(s). The crucial point is that, on this view, the kind of causality is the same whether the items causally connected are an intention and an action or just two billiard balls.[6]

The second assumption of the CTA is less often explicitly stated; nevertheless, I think it is as crucial as the first one. We can call this assumption a *uniform* understanding of content. On this view, there is only one kind of conceptual content that is uniform across *theoretical* and *practical* thought (e.g. belief and intention). Both kinds of thought are to be distinguished, not in terms of different kinds of content, but in terms of their different causal roles.[7] Very roughly, the basic idea is that, e.g. an intention is a *practical* thought in that it tends to cause what it represents, whereas a belief is a *theoretical* thought in that it is characteristically caused by what it represents – however mediated this causal dependence might be (see, e.g. Smith [1987, 55] for a clear statement of this view). According to this view, this distinction in terms of causal roles does not imply any difference in the respective kinds of content. So, given this view of content, it makes no intrinsic difference to the content of my practical thought that it is the content of a specifically *practical* thought, i.e. a thought that tends to cause what it represents.

Now, once such uniform notions of content and causality are in place, the CTA does indeed become the inevitable way of accounting for Davidson's insight. For, given these notions, we have to conceive of an intention's conceptual content and its causal role as two intrinsically independent components. But, I think we are now also in a position to better appreciate why the CTA is neither identical with nor entailed by Davidson's insight that, in acting intentionally,

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

one's intention is the cause of what it represents. For, just as such, this latter idea doesn't seem to involve or depend on any commitment to a uniform understanding of content and causality. Recall Davidson's argument for the claim that reasons – or, in our version, intentions – are causes: to exclude the possibility that one's movement conforms to one's intention by mere *accident*, one's intention must be the cause of the movement it represents, in the sense that it explains the *existence* of this movement. But this argument by itself doesn't imply a uniform understanding of causality; that is, it doesn't imply that the causal relation binding one's movement to one's intention is *not* intrinsically different from the causal relation that, say, binds fire to lightening. If this is right, one can accept Davidson's argument and the need to conceive of intentional action along the lines of (Intention) without *thereby* being committed to the CTA.[8]

Now, what I will argue in the rest of this paper is that it is not only necessary to distinguish between (Intention) and the CTA's way of accounting for it, but that given the CTA's understanding of (Intention), it actually becomes quite unclear how we can account for (Non-Accidentality). This, then, will motivate the exploration of a different way of understanding (Intention), one that rejects the CTA's underlying assumptions.

3. A problem for the CTA

3.1. A problem for the CTA's way of accounting for (Intention)

To bring out the difficulty facing the CTA, consider again that, on its view, in acting intentionally, an agent's intention must *both* represent *and* cause her movement, where both of these facts are intrinsically independent of one another. That is to say, considered as such, there is nothing in the notion of causality that requires the presence of representation; and, correspondingly, there is nothing in the notion of representation that requires the presence of causality. But, if that's so, it seems there is nothing that explains why it's *not* just an *accident* if representation and causality come together in the right way, namely so that what the intention causes is the movement it represents. In a case where the intention to do A causes one to do A, both conditions of the CTA are simultaneously satisfied: the intention *both* represents *and* causes one's doing A. But, given that content and causality are intrinsically independent of one another, how can this be anything but a *mere fluke*?

That this really is a problem can be illustrated by Davidson's (1973, 79) famous example of the nervous climber. In this example, a climber wants to rid himself of the weight and danger of holding another man on a rope, and believes that loosening his hold on the rope will accomplish this. As a result, he forms the intention to loosen his hold, which, however, makes him so nervous that he loosens his hold unintentionally. In this case, the intention both represents and causes the climber's loosening his hold, yet no intentional action is taking place. Why not?

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

Arguably, this is so because it is a mere *accident* that the climber's intention causes the movement it represents (or, equivalently, that it represents what it causes). For, in this case, the intention's causality operates independently of the fact that it represents its effect. To make this more vivid, suppose the climber's intention had been different – say, he had intended, not to loosen his hold, but to cut the rope on which his partner dangles. Presumably, this would not have made a difference to the resulting behavior. Given the climber's nervousness, it is plausible to assume that *both* intentions would have had the *same* effect, namely the climbers loosening his hold. This means that, under the circumstances, it is not the content of the climber's intention that is relevant to his behavior. And in this sense it is just an accident that the behavior caused by the climber's intention accords with the intention's content. So, what the example helps to bring out is that, even though both conditions of the CTA are satisfied, it may still be an accident that the agent's intention causes what it represents – and in that case, no intentional action is taking place.

Now, one might think this is a marginal case, one that can be put aside in the CTA's pursuit of an account of the standard case. But this would miss the point of the example. For, given the CTA's assumption that an intention's content and its causality are independent of one another, the problem is that nothing seems to rule out the possibility that in any given case things are just like in Davidson's example, namely that the agent's intention causes what it represents purely by accident. Thus, what is called into question is precisely the CTA's entitlement to distinguish in this way between 'marginal' and 'standard' cases. From the perspective of the CTA, any case looks like one in which it is just mere luck that the intention's effect coincides with what the intention represents.

It thus starts to look quite puzzling how, given the CTA's way of understanding (Intention), an intention's causality is supposed to account for (Non-Accidentality). For, if it really is just an accident that one's intention causes the movement it represents – and, on the CTA's view, it is hard to see how this can be otherwise – then, even though one's intention is the cause of one's movement, this doesn't ensure a non-accidental match between this movement and the content of one's intention. In Davidson's example, it is just an accident that the climber ends up doing what he intended to do – even though his movement is caused by his intention. But this means that, on the CTA's account of (Intention), the intention's causal role seems insufficient to account for the fact that, in acting intentionally, it is no accident that one is doing what one intends to do.

We may put this by saying that, strictly speaking, there are two sources of accidentality and, thus, two ways of violating the non-accidentality requirement on intentional action. First, this requirement is violated if the cause of the 'matching' behavior is something other than one's intention. But, second, this requirement is also violated if the causality of one's intention operates independently of its content. A proper account of the non-accidentality requirement

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

needs to exclude both aspects of accidentality. However, this is not to suggest that they can be dealt with separately. For, as we have just seen, even if the cause of the 'matching' behavior is the intention itself, this doesn't render the 'match' any less accidental as long as one hasn't also eliminated the second source of accidentality. That is, a proper account of the non-accidental 'match' between movement and intention requires an account of the non-accidental connection between the intention's causality and its content.

So, I think what we can conclude is this. When one acts intentionally, it is no accident that one is doing what one intends to do. But for this to be no accident, it is not enough that one's intention *both* represents *and* causes what one is doing. Both, the fact that the intention is a representation of one's movement and the fact that it is the cause of that movement cannot be *wholly* independent of one another. Rather, there must be some sort of connection between these facts, such that it is no accident if one's intention causes what it represents. The challenge can be further clarified by relating it to the problem of causal deviance – of which Davidson's nervous climber is usually taken to be the prime example.

On the present interpretation, what's going wrong in the example is that it is just an accident that the intention causes what it represents. Thus, what the example illustrates is the need for an account of the non-accidental connection between the intention's causality and its content. On my view, then, the problem of causal deviance is just an example of a more general problem: the problem of how to account for the non-accidentality requirement on intentional action. Now, if this diagnosis of the problem is correct, then it seems that many attempts to solve it are misguided from the very beginning: namely those attempts that fail to address the question of how causality and content must be *connected* in order to meet the non-accidentality requirement. Consider, for example, the so-called *causal immediacy* strategy (see, e.g. Brand [1984, 20] and Mele [1992, 202]). The general idea behind this strategy is that solving the problem requires eliminating potentially 'deviant' events (like nervousness or excitement) from the causal chain linking behavior to intention. And it is claimed that this can be achieved by stipulating that the intention *proximately* causes the behavior, without any mediating causal links. Now, setting aside any question as to whether this strategy can be successful in its own right, it should be clear that it doesn't help with our problem. For, even assuming that it succeeds in eliminating nervousness from the causal chain, this by itself does nothing to show that it is not just an accident that, in addition to being its cause, the intention also represents the behavior. What we are interested in is not how the causal chain is made up, but how it is related to the fact that the intention is a representation of the behavior.[9] In other words, what we need is not a specification of either the causal or the representational condition, but a specification of how they are connected, such that their *co*-presence is no accident.

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

3.2. Ways of solving the problem

Once we have a clear understanding of our problem, I think three ways of solving it suggest themselves, depending on how the non-accidental connection between causality and content is spelled out: the connection may be conceived of in *modal*, *functional* (or *teleological*), or *constitutive* terms. Corresponding to each of these accounts will be a different conception of accident. On the modal account, the relevant sort of accident is excluded if and because the agent is responsive to the content of her intention, not just in the actual situation, but also in a specified range of *non-actual* or *counterfactual* situations. Alternatively, it may be argued that it is no accident that the intention causes what it represents if and because doing so is its *function*, where this means that the intention's causal efficacy is in part explainable by the fact that it tends to bring about what it represents. Finally, the non-accidental connection between content and causality may be understood in *constitutive* terms, such that an intention is a *kind* of cause to which it is essential that it is a representation of its effect. Clearly, the constitutive account is incompatible with the CTA: it amounts to a rejection of the assumption that content and causality are intrinsically independent of one another. By contrast, I take the first two accounts to be compatible with the CTA. This is rather obvious for the modal account. After all, on this view, whether it is an accident that the intention causes what it represents is ultimately just a matter of the frequency with which it does so in a specified range of possible situations. It may be less obvious for the functional account. But here, everything depends on how the notion of function is itself understood, and I take it that a naturalistic understanding of this notion – one that appeals to natural selection – is indeed compatible with the CTA's commitments.

May there be other ways of spelling out the relevant non-accidental connection? Some have suggested that an intention must be causally efficacious *in virtue of* its content (see, e.g. Schlosser 2007). However, I think that, as such, this is just another way of saying that there must be a non-accidental connection between causality and content. If the 'in-virtue-of' relation is not just a placeholder – much like the appeal to being caused 'in the right way' – it must somehow be fleshed out. And, here again, modal, functional, and constitutive considerations suggest themselves as the most obvious candidates. Thus, in the end, I think that an appeal to a grounding relation such as 'in virtue of' doesn't get us anywhere we haven't already been with the options we are going to consider.

So, I think that at the heart of the issue is the question of how to understand the relevant sort of accident. In what follows, I will argue for the constitutive account of accidentality by criticizing the modal as well as the functional strategy. Consequently, if I am right, there is good reason to think that the non-accidentality requirement cannot be accounted for within the framework of the CTA.[10]

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

4. The modal account of accidentality

In more detail, the modal account of accidentality says roughly this: it is no accident that one's intention causes the movement it represents if, and only if, in a relevant range of counterfactual situations, one is responsive to one's intention. Depending on which counterfactual situations are considered relevant, two ways of spelling out the modal condition suggest themselves. First, one might consider situations in which the agent's intention *differs* from the one in the actual situation. Here, if the agent is responsive to what she intends, we would expect her behavior to differ correspondingly. Alternatively, one might consider situations in which the agent's intention is the *same* as in the actual situation. Here, if the agent manifests the relevant sort of responsiveness, we would expect her behavior to be the same as well. The basic idea is that supplementing the two core conditions of the CTA – the representational and the causal condition – with one of these modal conditions will secure the relevant sort of non-accidentality.[11]

4.1. Modal condition I

The first version of the modal condition corresponds to what has been called the *sensitivity strategy*.[12] It requires that the agent's behavior is responsive to the content of the intention that causes the behavior in the following sense: if the content of the agent's intention had been different, the resulting behavior would have differed correspondingly. If the agent satisfies this condition, it is *no* accident that the caused behavior conforms to the content of her intention. For instance, Davidson's climber does not meet this condition, because, as we have seen, it is *not* true that, if the content of his intention had been different, the resulting behavior would have been different too. However, as it stands, the condition is clearly too strong. For, in order to act intentionally, it is certainly not necessary that in just *every* counterfactual situation in which one had intended something else, one would have acted in conformity with this different intention. After all, innumerable things may happen and prevent one from responding successfully to one's intention. Consider a different version of Davidson's example. Suppose the climber isn't nervous at all; instead, he goes through with his intention to loosen the hold on the rope in a very cold-blooded manner. Now, suppose further that, if his intention had been different – e.g. if he had intended to pull up the rope and save his companion – some external force (an evil demon, a falling rock, a sudden heart attack) would have prevented the climber from acting on this different intention. In a case like this, the relevant counterfactual is false: it is *not* true that, if the content of the climber's intention had been different, his behavior would have differed correspondingly. Nevertheless, this doesn't show that, in the actual situation, the climber didn't act intentionally. Hence, at least as it stands, the sensitivity condition doesn't seem necessary for intentional action.[13] But,

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

since the relevant sort of non-accidentality *is* necessary for intentional action, sensitivity seems inadequate to account for it.

Still, there seems something deeply right about the sensitivity condition. Normally, when someone acts intentionally, we *do* expect that, if she had intended something else, her behavior would have differed accordingly. It's just that we don't expect this *no matter what*. We only expect this in circumstances where nothing interferes with the agent's ability to act on her intention. Thus, clearly, if the relevant counterfactual stands a chance of being true, its scope must be restricted to circumstances in which possible interveners are absent. Yet, the question is how to accomplish this restriction. It seems hopeless to produce a list that includes all possible interveners. Since there is no obvious limit to the things that might interfere with the agent's ability to carry out her intention, such a list would likely be endless. Certainly, the more natural move is to amend the counterfactual with a *ceteris paribus* clause. However, it is not clear that this move is compatible with the reductive ambition of the CTA. For, as is well known, the introduction of such a clause gives rise to the problem that it seems impossible to understand the clause in a non-circular way.

To see this, consider the events that would have to be excluded by the *ceteris paribus* clause. Among those events are falling rocks, heart attacks, and evil demons. Now, ask yourself, what do these events have in common with each other such that they are all covered by the *ceteris paribus* clause? It seems difficult to answer this question other than by saying something like: 'each of these events prevents the agent from properly realizing her intention'. Thus, what justifies grouping these events together can be specified in only *negative* terms: they all interfere with the agent's ability to act intentionally. This is the respect in which events must resemble one another if they are to be covered by the *ceteris paribus* clause. But, if that's so, it is impossible to specify the relevant respect of similarity without making reference to the notion of intentional action itself. An understanding of the *ceteris paribus* clause would thus contain and depend upon an understanding of acting intentionally. This means that, if we render the relevant counterfactual true by introducing a *ceteris paribus* clause, we thereby render it unavailable to a reductive account of intentional action. As I said, I don't want to deny that a counterfactual of the relevant sort is true when one acts intentionally. What I claim is just that a proper understanding of the counterfactual is not more fundamental than the notion of intentional action itself.

Arguably, there are other, perhaps more sophisticated versions of the sensitivity strategy. Recently, for example, it has become popular to claim that, in acting intentionally, an intention must not just trigger but continue to *guide* a movement through the course of its occurrence.[14] It is tempting to understand this guidance-condition as an attempt to account for the non-accidentality requirement. Thus, for it to be no accident that one is doing what one intends to do, one's intention must cause one's movement in such a way that it is *guided* through the course of its occurrence. But what does 'guidance' mean

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

here? A natural interpretation appeals to a certain sort of counterfactual: one's movement was guided in the relevant sense just in case, had the successful completion of one's movement been in danger, there would have been a corresponding correction of one's behavior.[15] Suppose, for instance, in letting go of the rope, Davidson's climber was in fact guided by his intention to do so. Then, a counterfactual of the following sort is true: had he encountered an obstacle – e.g. had the rope been stuck – he would have corrected or modified his behavior accordingly, e.g. he would have released the rope.

The problem with this account is the same as the one we just discussed. As it stands, the relevant counterfactual is too strong. Imagine that, in the actual situation, the climber didn't encounter any obstacles, but if he had, some external force would have prevented him from correcting his behavior. In this case, the counterfactual is false, but this does not imply that, in the actual situation, he didn't act intentionally. So, to save the account from counterexamples, one has to amend the counterfactual with a *ceteris paribus* clause. But then, again, the account threatens to become circular. For, it is difficult to see how the set of events covered by the *ceteris paribus* clause can be specified other than in *negative* terms: as those events that interfere with an agent's intentional agency.

4.2. Modal condition II

Now, let's turn to the second version of the modal condition we distinguished above – the one where the agent has the *same* intention in the relevant non-actual situations as in the actual situation. In spelling out this condition, we can draw on some recently influential work in epistemology, according to which the notion of an accident is understood in terms of possible worlds. In the words of Duncan Pritchard (2005, 130–31), the basic idea is this: 'when we describe an event as an accident we thereby typically imply that in a wide class of the relevant nearby possible worlds it doesn't occur'. This is a general claim about what constitutes an accident. If we adopt this general view to our case, we arrive at the following characterization: it is an accident that the agent's intention to do A causes what it represents if, and only if, in most nearby possible worlds where the agent intends to do A, the intention does *not* cause her to do A and, thus, she is *not* doing A. So, if we want to exclude this sort of accident, we will have to amend the CTA with the following modal condition: S is doing A intentionally only if, in most nearby possible worlds where S intends to do A, S is doing A.[16]

Just like with the previous proposal, I think that, as it stands, the condition is too strong, and that an improvement on the condition is likely to render it unavailable to a reductive account of intentional action. To begin, consider what is meant by 'nearby' in 'nearby possible worlds'. The qualification is intended to restrict the set of possible worlds relevant to the assessment of whether what is happening in the actual world is an accident. Thus, importantly, to determine

whether what's going on in the actual world is an accident, we don't need to look to just *any* possible world. We only need to look to possible worlds that are sufficiently *similar* to the actual world (or, as it is sometimes put, to possible worlds that could have easily been the case). However, I think the task of specifying the right degree and respect of similarity between worlds gives rise to difficulties very much like those we encountered in specifying the content of the *ceteris paribus* clause. In other words, it strikes me as quite unlikely that we can specify how similar a possible world must be to qualify as 'nearby' in the relevant sense, without making reference to the notion of intentional action.

To bring this out, consider the following case. John intends to kill a tyrant by poisoning the supply of drinking water to the tyrant's palace. However, unbeknownst to John, the tyrant is well aware of his plans and orders Jones to prevent John from realizing his intention. Now, suppose further that, due to some quite unlikely event, Jones himself is prevented from interfering with John's plans – imagine, for example, that he is run over by a car on his way to where John intends to poison the water supply system. In that case, John will realize his intention without any interference by Jones. But there are many nearby possible worlds in which John is *not* doing what he intends to do, i.e. worlds in which Jones is not run over by a car and prevents John from poisoning the water supply. Thus, the relevant modal condition is false: in many nearby possible worlds where John intends to poison the water supply, he is *not* doing so. However, this doesn't seem to imply that, in the actual case, John is not acting intentionally when he is poisoning the water supply. Hence, as it stands, the relevant modal condition doesn't seem necessary for intentional action and, thus, it is inadequate to account for the non-accidentality requirement.[17]

In response, one might want to argue that, although worlds in which Jones is not run over by a car and prevents John from poisoning the water supply are indeed possible and nearby in some sense, they are not sufficiently similar in the *relevant sense*. To determine whether, in the actual case, John is acting intentionally, we should consider only those worlds in which Jones is run over by a car. If we do, the relevant modal condition may turn out to be true after all. Perhaps, this is the right way to preserve what seems intuitively right about the modal condition. But if that's the way to save the condition from counterexamples, it is hard to see how the relevant respect of similarity can be specified without reference to the notion of intentional action. It certainly seems that, when we exclude worlds in which Jones is not run over by a car from the set of worlds that qualify as relevantly similar, we rely on a previous decision to treat what John actually did as an intentional action. Thus, in ascertaining whether the similarity to a nearby possibility is relevant or great enough, we do seem to be guided by our grasp of what it is to do something intentionally.[18] Moreover, examples such as the one we just considered can be easily multiplied. But if that's so, then things start to look quite the opposite from what the modal account suggests: it is not that we determine whether

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

someone is acting intentionally by considering if the relevant modal facts obtain, rather we determine whether the relevant modal facts obtain by considering if someone is acting intentionally. Once again, I don't want to deny that, when someone is acting intentionally, a modal condition of the relevant sort is true. It's just that I don't think the condition is intelligible independently from the idea of intentional agency itself, which it would have to be if it is to be employed in a reductive account of this idea.

So, to sum up, if these considerations are correct, the attempt to properly understand both versions of the modal condition returns us to the notion of intentional action. This is to say, even though it is true that, in acting intentionally, an agent is modally responsive to her intention, it is not *because* she is modally responsive that she is acting intentionally. Rather, the connection between intentional action and modal responsiveness runs in the other direction: it is *because* an agent is acting intentionally that the relevant modal condition is true. Now, what matters for our purposes is that this suggests that there should be a *non*-modal account of the non-accidentality requirement. For, if a proper understanding of the modal condition is dependent on the notion of intentional action, then by the same token it is dependent on a notion of non-accidentality, the one that is implied by the notion of intentional action. Hence, if we don't want to rest content with a *brute* notion of non-accidentality, we should be looking for a *non*-modal account of the non-accidental connection between an intention's causality and its content. Recall that earlier I distinguished between three different ways of accounting for this connection: in modal, functional, or constitutive terms. In the next section I will briefly discuss and dismiss the functional account. If this is right, the need for a non-modal account suggests that we should conceive of the relevant non-accidental connection in constitutive terms – something I will turn to in Section 6.

5. The functional account of accidentality

Responding to the problem of causal deviance, Enç (2003, 108) claims that the 'key to understanding what counts as a "normal" causal chain is the concept of executing a function'.[19] The appeal to the notion of function is relevant to us because it offers a distinctive account of the non-accidental connection between causality and content. The idea, very roughly, is that it is no accident that the intention causes what it represents if and because doing so is the intention's function, where this implies that the intention's causal efficacy is somehow explainable by the fact that it tends to bring about the intended end. Thus, on this view, the CTA is to be supplemented by a condition specifying the explanatory dependence of the intention's causality on its tendency to bring about what it represents.

Central to the functional account is a distinction between effects of an intention that manifest its function and those that do not. Crucially, this difference is not to be spelled out in merely dispositional or reliabilist terms. To say

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

that it is an intention's function to cause what it represents means more than to say that the agent is reliably disposed to respond to her intention by doing what it represents. Consider the heart's function of pumping blood around the body. Apart from pumping blood, a heart does many things, some of them very reliably (such as making a thumping noise). But only pumping blood is an execution of the heart's function. So what's the difference? The traditional thought of course is that the attribution of functions implies a certain kind of explanation: pumping blood is the heart's function because it figures in an explanation of why the heart does what it does to achieve pumping blood. Putting this in more general terms, what is characteristic of functional explanations is that we can explain why an object does something by the fact that doing so tends to achieve a certain result. Now, in order to spell out this notion of an explanatory dependence, contemporary accounts of function typically turn to the theory of *natural selection*.[20] Very roughly, the idea is that the heart's function is pumping blood because doing so conferred an evolutionary advantage to certain organisms (say, humans) in that they were more likely to survive and reproduce than their competitors with different traits. In that sense, an organism's heart owes its existence to what hearts did (i.e. pumping blood) in that organism's ancestors. Pumping blood can thus be said to figure in an explanation of why hearts do what they do in the sense that pumping blood is something hearts did in the past that explains the existence of hearts and their dispositions in the present.

Now, if we apply this etiological model of biological functions to intentional action – as Enç (2003, 105–115) does – we arrive at roughly the following view. As noted, reliability of effect is not enough for attributing a function to an intention. What is more is that it involves an account of why the intention is reliably efficacious in this way, namely one that points to its tendency to bring about what it represents. In that sense, an intention can be said to be causally efficacious for the sake of producing the intended end. Given the etiological model, what ultimately underwrites the explanatory dependence of the intention's causal efficacy on its tendency to realize its content is some story about natural selection. Again, putting it crudely, this means that reliably responding to an intention by doing what it represents has contributed to the survival and reproduction of human agents in the past, which is why this is what we do in the present. In other words, bringing about what it represents is an intention's function because this is the effect for which it was selected by evolution.[21] However, as I shall argue now, if we conceive of an intention's function on this biological model, the attribution of function seems neither necessary nor sufficient for intentional action.

First of all, it is not clear how this account applies to intentions whose objects are *not* actions relevant to survival and reproduction (such as eating, drinking, and sex). Think of my intention to buy the latest iPhone model: no story about natural selection seems available to back up the attribution of a function. But even if we bracket this concern and assume that the account

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

can somehow be extended to intentions at large, there is a problem for the sufficiency of this proposal. For, as Mayr (2011, 132) has objected against Enç, '[t]he biological function of intention-states is clearly not limited to producing things over which we have agential control.' That is, there are causal effects of an intention that manifest their biological function, but which are clearly not intentional actions. Mayr's example is the intention to eat which usually brings about certain physiological changes – such as the production of saliva – necessary for the realization of this intention. If this is right, we can imagine a case in which an agent not only intends to eat, but also intends to produce saliva at the same time (perhaps he doesn't know that the production of saliva is not something he can do intentionally). If, as a result, saliva is produced, the intention's effect conforms to its content and it was caused 'in the right way', because the production of saliva is a manifestation of the intention's biological function. However, despite all this, producing saliva is not something that the agent is doing intentionally.[22] So, in this case, even though we may want to say that there is some sense in which it is no accident that the intention causes what it represents – namely saliva – this is clearly not the sort of non-accidentality we are after. This is brought out by the fact that the intention would have produced saliva, even if doing so had not been part of the intention's content. But, as we have seen in the previous section, the truth of such modal facts *is* implied by the relevant sort of non-accidentality – it's just that these facts are not fundamental in that they themselves depend on a non-modal account of accidentality.

6. The constitutive account of accidentality

Both the modal and the functional strategy seek to capture the relevant sort of accident by looking beyond the agent's present situation: while the modal account is looking at what *might* or *would* have happened in certain counterfactual situations, the functional account is looking at what happened in the agent's (evolutionary) *past*. By contrast, on the constitutive account, what is needed is a better understanding of the kind of causality *actually* at work when someone is acting intentionally. More specifically, what is required is an understanding of the intention's causality such that its explanatory power is itself *sufficient* to exclude the relevant sort of accident. And the basic idea is that this can be achieved if we conceive of the connection between causality and content in *constitutive* terms, such that an intention is a kind of cause to which it is essential that it is a representation of its effect. In this section, I will present a rough outline of what this claim amounts to. However, a fully fledged defense of the constitutive account is beyond the scope of the present reflection.[23] My aim here is more limited: given a certain view of the causality involved in intentional action, I argue that this view can be deployed in an account of the relevant sort of non-accidentality.

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

Let's begin then with considering, on the most abstract level, what it means to claim that the intention's representation is essential to its causality. In the most general terms, it means that there is a *distinctive* kind of causality, one that is crucially different from the kind of causality connecting items that do *not* stand in a representational relation to one another (like the notorious billiard balls). In other words, the constitutive account rejects the CTA's assumption of a *uniform* notion of causality, according to which causality is everywhere the same. Instead it requires a more pluralistic take on causality: it requires that we conceive of the notion of causality as a category that allows for different specifications. On this view, the sort of causality underlying a true action-explanation is different from the one making true an explanation such as 'The house is burning down because it was hit by lightning'. What justifies speaking of *causal* explanations in both cases nonetheless is that each one provides an account of the *existence* or *actuality* of the explanandum. Thus, roughly, what unites the different species of causality under a single category is that each one is a real relation that can be revealed in an explanation telling us why the world is the way it is in a certain respect. Such a pluralistic approach to causality has gained increasing support in recent years; defending it, however, would require (at least) another paper.[24] Again, my strategy here is different: assuming that, in principle, it is possible to distinguish between different kinds of causality, I shall consider what makes the sort of causality involved in intentional action special, and how its distinctive character can help us capture the relevant sort of non-accidentality.

To do this, I think it is helpful to turn to Elizabeth Anscombe's (2000) approach to understanding intentional action. Famously, on her view, intentional action is to be understood through the specification of 'a certain sense of the question "Why?"' (2000, 9). Thus, she defines intentional action as a movement that is subject to a certain kind of explanation. And, in fact, I think that her investigation of the relevant question 'Why?' is best understood as the project of specifying a distinctive kind of causality – the one revealed in the relevant kind of explanation.[25] Thus, in sharp contrast to the CTA, Anscombe doesn't apply a preconceived notion of cause to the account of intentional action; rather, specifying a special kind of cause is itself an integral part of such an account. More specifically, what she takes to be distinctive of the relevant form of explanation – the one that defines intentional action – is its essentially *self-conscious* character. At one point she puts this by saying that the relevant 'Why?' – question 'is *refused application* by the answer: "I was not aware I was doing that"' (2000, 11; my italics). Thus, importantly, her claim is *not* that there is one kind of 'because', which is an explanation of something I am doing intentionally if, in addition to being true, I am also aware of it. Rather, her claim is that there is a kind of 'because' that applies – is true – only if I am aware of it. Thus, the appeal to the agent's awareness singles out a distinct kind of 'because', it doesn't just identify conditions under which an otherwise understood 'because' yields an explanation of intentional action.

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

Roughly, what this means is that, for the relevant explanation to be true, the acting subject herself must be able to give it. For example, when we say 'John is cutting tomatoes because he intends to make a gazpacho', and this is a true explanation of the relevant kind, this necessarily requires that John himself can answer the question of why he is cutting tomatoes by citing his intention to make a gazpacho. In the explanation, the 'because' designates a causal connection that wouldn't obtain if John didn't understand himself as realizing the intention to make a gazpacho.[26] Usually, if one is realizing one's intention to make a gazpacho, one is (already) in the process of making a gazpacho – something expressed by the progressive action-description 'I am making a gazpacho'.[27] Thus, in understanding himself as realizing the intention to make a gazpacho, John must conceive of what he is doing under the description provided by his intention: he must be able to think 'I am making a gazpacho'. We may put this by saying that what Anscombe's investigation of the question 'Why?' reveals is a special kind of cause: an intention's causal power is such that it explains what one is doing in a way that includes the agent's understanding of her doing as realizing her intention (and, thus, the representation of her doing under the description provided by the intention).

To be sure, Anscombe's view of intentional action is itself controversial, and this is not the place to mount a proper defense of this view. My point here is just that, on the face of it, it provides us with the material for an account of the non-accidentality requirement, and, given that the modal and the functional account seem insufficient, there is reason to think that a proper account of this requirement may indeed turn on our ability to make sense of this view.

So, assuming Anscombe's view is correct, how does it provide us with a sense in which it is no accident that an intention causes what it represents? To see this consider that, on this view, the relevant causal relation is partly constituted by the fact that the agent represents her behavior under the description provided by the intention – in that sense the intention is a kind of cause to which it is essential that it is a representation of its effect. And if that's so, then, surely, it is no accident that the intention causes what it represents. That is to say, on the Anscombian characterization of an intention's causality, the intention's causal power all but guarantees that its effect matches its content. Thus, I think that the idea of a distinctive kind of self-conscious causality would indeed help us capture the relevant sort of non-accidentality. It puts us in a position to account for the non-accidentality requirement in primarily *explanatory* terms: in acting intentionally, it is no accident that one is doing what one intends to do if and because what one is doing is subject to a *special kind of causal explanation*, one whose truth essentially involves the agent's understanding of this explanation. By contrast, if it *is* just an accident that one is doing what one intends to do, then no such explanation is available. To illustrate, consider once again Davidson's nervous climber. On the present view, it is just an accident that the climber ends up doing what he intends to do because, in this case, the causal nexus linking the climber's behavior to his

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

intention is of the *wrong kind* – for it doesn't essentially involve the climber's understanding of this nexus.[28] And, as the case is described, this seems independently plausible: in letting go of the rope, the climber may or may not be aware of what's going on; but whether or not he is thus aware doesn't affect the causal nexus, which obtains regardless of what the climber thinks.

Thus, on the present view, fulfillment of the non-accidentality requirement ultimately turns on the availability of a certain kind of causal explanation. This, I think, allows us to understand the causal power of intentions in such a way that it is indeed *sufficient* to account for the relevant sort of non-accidentality. Recall that, according to Davidson's original argument, for it to be no accident that one is doing what one intends to do, one's intention must be the cause of what it represents. As we have seen, given the CTA's understanding of this idea, the intention's causality is insufficient to ensure a non-accidental match between movement and intention. This is why proponents of the CTA turn to additional – especially, modal – conditions to secure the relevant sort of non-accidentality. But, on the present view, what drives the search for an additional condition is just an *impoverished* conception of causality. For, on the present understanding of an intention's causality, its explanatory power is in fact sufficient to account for the non-accidentality requirement. There is no need to look elsewhere for additional conditions. A case like the one of Davidson's climber is excluded because it involves the wrong kind of causality, not because the climber fails to meet some additional *non*-causal condition. Again, this is not to deny that, if one is acting intentionally, certain modal facts will obtain. It's just that these modal facts are not fundamental: they reflect underlying truths about the explanatory relationship between intention and action.

Against the present view, one might want to object that representation cannot be constitutive for causality because an intention may fail to cause what it represents. Presumably, this is shown by the possibility of making mistakes, as when I intend to walk to the beach, but took a wrong turn at some point.[29] However, I think proponents of the constitutive account can reply in either of two ways. First of all, it may be questioned whether the objection relies on a correct description of cases of making a mistake, namely as cases where one doesn't count as being engaged in the action intended at all. As has often been noted, progressive action-descriptions ('I am walking to the beach') don't imply success (having been hit by a car, I might never make it to the beach), nor do they imply that, at this very moment, I am doing anything that moves me closer to my end (as when I am waiting at a traffic light).[30] Just as well, it seems important that their truth is compatible with making mistakes (such as taking a wrong turn), at least as long as one is willing to correct them. For, otherwise, there would be no room for the distinction between doing something *badly* and *not* doing it at all. But, after all, success in acting intentionally is a matter of degree, not of all or nothing. Taking this into account, it might be argued that, properly understood, making a mistake doesn't entail that one

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

isn't engaged in doing what one intends to do at all, but rather that it qualifies the *way in which* one is doing it.[31]

Second, even granting that causality and representation *can* come apart – as in cases of making a mistake – this doesn't show that they *must* come apart. When an intention's effect fails to realize what it represents, this doesn't have to be understood as a case where, though the intention's causality is working properly, the agent fails to meet an additional representational condition. Instead, and I think more plausibly, we can understand such a case as one where the exercise of the intention's causal power is itself *defective* or *imperfect*. On this view, an intention's causal power is *defined* by cases of its successful exercise, i.e. by cases where it is productive of the movement represented. Cases where the intention is *not* productive of the movement represented must be understood as cases where something interfered with the proper exercise of an intention's causal power (e.g. something that distracted the agent). The point is that successful actions can usually be explained by the agent's intention *alone*, whereas the explanation of unsuccessful actions must include reference to an *additional* element, an interfering factor driving causality and representation apart.[32] Contrast this with the CTA: when an intention isn't productive of the movement represented, then by itself this doesn't imply that something must have interfered driving causality and representation apart – for, on the CTA, causality and representation have been apart all along. Instead, it is the successful case that calls for *extra* explanation – the case where, not by accident, the intention causes what it represents. For, as we have seen, to account for the non-accidental *co*-presence of causality and representation, we need to invoke some *additional* – e.g. modal – condition.

To be sure, these remarks leave many questions open as to how an account of intentional action along the lines of the constitutive view is properly developed and defended.[33] But, again, my aim is not to present a fully fledged account of this sort. Rather, it is to suggest that this is what is needed if we want to account for the non-accidentality requirement on intentional action.

Acknowledgments

For very helpful discussions or comments, I am grateful to Hagit Benbaji, David Enoch, Matthias Haase, David Hunter, Andrea Kern, Sebastian Rödl, Karl Schafer, Sergio Tenenbaum, and two anonymous referees for this journal. Thanks also to audiences at the universities of Jerusalem, Beer Sheva, Tel Aviv, Leipzig, and Utrecht.

Notes

1. We can replace 'reason' with 'intention' in the sense that the argument's core does not depend on Davidson's view that acting intentionally is necessarily acting for a reason, as opposed to merely acting on an intention. For, even if the necessity of reasons is denied (as it is by, e.g. Anscombe 2000; Setiya 2007), we can run the same argument: to rule out that the agent's movement accords with her intention by mere accident, the intention must be the cause of the movement.

2. Here and elsewhere I am assuming that, at least in the most basic cases, an intention represents one's movement. Sometimes it is said that an intention represents a (basic) action, not a movement. For present purposes, nothing much hangs on this. So one might as well replace 'movement' with 'action'.
3. Aside from Davidson himself, prominent examples of the CTA include Bishop (1989), Velleman (1989, 2000), Setiya (2007), Bratman (1987, 1999), Mele (1992), Searle (1983, 2001), Harman (1976), and, arguably, Frankfurt (1988).
4. Davidson's claim that, in acting intentionally, the agent represents her action as in some sense good or valuable is controversial (for influential criticism, see Setiya 2007; Velleman 2000). However, as it is usually framed, this controversy doesn't affect the basic structure of the CTA, since it doesn't change the assumption according to which representation (or rationalization, or justification) and causality are intrinsically independent of one another. This is why it will not be of concern to the argument in this paper.
5. In the quote, I have replaced 'event-causalist' with 'CTA' to preserve terminological continuity.
6. However, the uniform understanding of causality is by no means specific to Davidson. It cuts across theories of causality that may differ in otherwise important respects. For instance, it may be shared by non-reductive as well as reductive (e.g. regularity) theories of causality. For critical discussions of the uniform notion of causality, see, e.g. McDowell (1994), Steward (1997), Hornsby (2001), and Marcus (2012).
7. This view is reflected in the common understanding of mental states like intentions or beliefs as *propositional attitudes*. Implicit in this understanding is the assumption that the contents of theoretical as well as practical thoughts can be captured in terms of propositions to which different attitudes may be attached, attitudes which are usually distinguished in terms of their different causal roles.
8. By the same token, we have to be careful to distinguish between (Intention) and Davidson's particular view of causality. By itself, the argument for the claim that reasons are causes in no way implies or requires his nomological view of causality. Consequently, it is possible to accept the claim that reasons are causes without *thereby* being committed to the claim that actions and reasons must be identical with particular events instantiating a strict natural law. In fact, moving from the one to the other claim – as Davidson (1970) does in his argument for Anomalous Monism – crucially depends on the assumption of a uniform understanding of causality.
9. The same is true for attempts to solve the problem by further specifying the content of the agent's intention, e.g. by including an *action-plan* in the intention's content. (For a view of this sort, see, e.g. Harman 1976.) For, however precise the intention's content may be, this doesn't show that there is a non-accidental connection between the fact that the intention represents the behavior and the fact that it is its cause.
10. However, even though the considered options strike me as the most natural ways of accounting for the relevant sort of non-accidentality, I don't have a principled argument for why they *must* exhaust all possible option. Hence, the following argument for the constitutive account is limited in this respect: eliminating the rival options counts in favor of the constitutive account only insofar as there are no relevant alternatives.
11. In fact, I think the modal account covers a variety of accounts that, at first, might look rather different. Thus, for instance, I take it that the attempt to secure the relevant sort of non-accidentality by requiring that the agent's behavior be the output of a reliable disposition to respond to her intention's content is just a *version*

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

of the modal account (for a view of this sort, see, e.g. Brandom 1994, 235; 261). For, presumably, the most promising understanding of reliability is itself in modal terms. As we will see later, the same is true for accounts that appeal to the notion of guidance (as do, e.g. Raz 2011; Setiya 2007).

12. For an extended discussion of this strategy, see Bishop (1989, 148–75).
13. Responding to a similar counterexample, Bishop attempts to secure the necessity of the counterfactual by extending it to include a stepwise counterfactual dependence with a ban on backtracking counterfactuals (see Bishop 1989, 152). However, even if such an account could be defended against counterexamples, this move strikes me as completely ad hoc, since the sole motivation for supplementing the sensitivity condition in this way is that otherwise it would founder on counterexamples.
14. See, e.g. Setiya (2007, 32) and Raz (2011, 28, 31–32). The idea goes back to Frankfurt (1978).
15. Sometimes the notion of guidance is explicated in terms of a *negative feedback mechanism* (see, e.g. Bishop 1989, 167–72; Raz 2011, 31–32). But, ultimately, I don't think this is a different account. For, arguably, the reason why a feedback mechanism is relevant to the understanding of guidance is that it entitles us to supplement the CTA with some counterfactual or conditional condition specifying that, if the successful completion of one's action is jeopardized, it would be corrected or modified. For more on this, see Horst (2012, 58–77).
16. This condition is analogous to the so-called *safety* condition in epistemology, which is introduced to account for the idea that, if someone knows that p, it is no accident that her believe that p is true. Roughly, the safety condition is this: S knows that p only if, in most nearby possible worlds where S believes that p, p is true. Defenders of safety include notably Sosa (1999), Williamson (2000), and Pritchard (2005).
17. Counterexamples with a similar structure have been presented against the safety condition in epistemology. See Neta and Rohrbaugh (2004).
18. In this respect, I agree with an analogous claim by Williamson (2000, 100–1) concerning a safety condition in epistemology. He explicitly denies that this condition can be used in a non-circular analysis of knowledge, precisely because he doesn't think that we can specify the relevant degree and kind of similarity among possible worlds without using the concept of knowledge itself.
19. Enç credits Davies (1983) for noticing the importance of teleological considerations for solving the problem of causal deviance.
20. For prominent accounts of functions in terms of natural selection, see Millikan (1989) and Neander (1991). Here, I will just grant that this view is the correct account of biological functions (for criticism, see, e.g. Plantinga 1993, chap. 11), but question its application to intentional action. I focus on natural selection theories of function for two reasons: (a) such theories are uniquely suited to the CTA's aim to account for intentional action in terms of a uniform notion of causality; (b) rival theories can be ruled out, I think, because they are either irrelevant, subsumable under the modal account, or incompatible with the CTA. Thus, for instance, intentional design theories of function may be good for artifacts, but do not explain the function of intentions because human agents are not intentionally designed (moreover, there would be obvious worries of circularity). More plausible would be an appeal to causal role theories of function (see Cummins 1975 for a classic statement of this view). But such theories are better understood as versions of the modal account, since they explain functions in terms of dispositions. Finally, accounts that appeal to irreducible teleological notions can be ruled out here because they are incompatible with the CTA's aim to account for intentional

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

action in terms of a notion of causality applicable to inanimate objects (like billiard balls) as well.

21. In presenting the functional account, I have largely abstracted from the details of Enç's own version of this account, which is significantly more complex (see Enç 2003, 105–15). This is justified because the criticism of the functional account – to which I will turn now – doesn't depend on any particular details of the view.
22. For this counterexample, see Mayr (2011, 132–33).
23. But see, in particular, Rödl (2007), Marcus (2012), and Horst (2012) for a more detailed development and defense of the sort of view of intentional action that I'm going to sketch.
24. For recent attempts to explicate and defend this idea, see especially McDowell (1994), Steward (1997), Hornsby (2001), Rödl (2007), and Marcus (2012).
25. For reasons of space, I cannot defend this reading of Anscombe here. But compare, e.g. Rödl (2007), Marcus (2012), Ford (2011), Horst (2012), and Lavin (2013) for roughly congenial interpretations of *Intention*. Often, Anscombe is read as rejecting any role of the notion of cause in an account of intentional action. But I believe that a more sensible reading has her reject a *uniform* notion of causality, not just any notion of cause. At any rate, the systematic point of the argument doesn't depend on whether or not this reading of Anscombe is correct.
26. That a true action-explanation must involves the agent's understanding of that explanation doesn't mean that, in acting, she must consciously consider or reflect on the explanation. In general, understanding (just like knowing) P doesn't require that one is presently conscious of P. One may know that Madrid is the capital of Spain without being presently conscious of that fact. All it means is that, other things being equal, the relevant explanation is available to the agent's conscious reflection, such that she is in a position to answer the relevant question 'Why?'.
27. To clarify: I am not claiming that it is correct to say 'S is doing A' whenever S is doing something because he intends to do A. Rather, the claim is that whenever S is doing something intentionally there will be an intention with respect to which this is correct to say.
28. This means that an intention can be causally efficacious in a way that doesn't manifest its characteristic causal power, as when an intention causes one to be nervous or to blush. However, on the constitutive account, such cases are only possible because there are also other cases, namely cases where an intention explains one's action in a way that includes the agent's understanding of this explanation. These latter cases are fundamental in the sense that they are *definitive* of what it is to intend something in the first place; without them, there wouldn't be any other cases.
29. These cases are different from the ones we considered before – where an intention causes one to be nervous or to blush – because, presumably, even though the agent is making a mistake, what she is doing is still an exercise of her intentional agency.
30. See, e.g. Thompson (2008, Part Two) and Falvey (2000).
31. A view of this sort is developed and defended by Haase (Unpublished manuscript). See also Falvey (2000, 29).
32. For this sort of view, see Rödl (2007, 175–77) and Marcus (2012, 87–90).
33. For instance, as I have argued elsewhere (Horst 2012), I believe that a proper development of this view would also require a more explicit and substantial revision of the CTA's second basic assumption – the *uniform* understanding of content.

https://doi.org/10.1080/00455091.2015.1080045 Published online by Cambridge University Press

References

Anscombe, G. E. M. 2000. *Intention*. Cambridge, MA: Harvard University Press.
Bishop, John. 1983. "Agent-causation." *Mind* XCII: 61–79.
Bishop, John. 1989. *Natural Agency*. Cambridge: Cambridge University Press.
Brand, Myles. 1984. *Intending and Acting: Toward a Naturalized Action Theory*. Cambridge, MA: M.I.T. Press.
Brandom, Robert. 1994. *Making It Explicit*. Cambridge, MA: Harvard University Press.
Bratman, Michael. 1987. *Intentions, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
Bratman, Michael. 1999. *Faces of Intention*. Cambridge: Cambridge University Press.
Cummins, Robert. 1975. "Functional Analysis." *The Journal of Philosophy* 72 (20): 741–765.
Davidson, Donald. 1963. "Actions, Reasons, and Causes." Reprinted in *Essays on Actions and Events*. Oxford: Clarendon Press. 1980.
Davidson, Donald. 1970. "Mental Events." In Experience and Theory, Reprinted in *Essays on Actions and Events*. edited by Lawrence, Foster, and J. W. Swanson, London: Duckworth. 1980.
Davidson, Donald. 1973. "Freedom to Act." In Essays on Freedom of Action, Reprinted in *Essays on Actions and Events*. edited by Ted, Honderich, London: Routledge. 1980.
Davidson, Donald. 1982. "Paradoxes of Irrationality." In Philosophical Essays on Freud, Reprinted in *Problems of Rationality*. edited by R. Wollheim, and J. Hopkins, Cambridge: Cambridge University Press. 2004.
Davies, Martin. 1983. "Function in Perception." *Australasian Journal of Philosophy* 61: 409–426.
Enç, Berent. 2003. *How We Act*. Oxford: Oxford University Press.
Falvey, Kevin. 2000. "Knowledge in Intention." *Philosophical Studies* 99 (1): 21–44.
Ford, Anton. 2011. "Action and Generality." In *Essays on Anscombe's "Intention"*, edited by Anton Ford, Jennifer Hornsby, and Frederick Stoutland, 76–104. Cambridge, MA: Harvard University Press.
Frankfurt, Harry. 1978. "The Problem of Action." Reprinted in *The Importance of What We Care About*. Cambridge: Cambridge University Press.
Frankfurt, Harry. 1988. *The Importance of What We Care about*. Cambridge: Cambridge University Press.
Haase, Matthias. "Knowledge and Error in Action." Unpublished manuscript.
Harman, Gilbert. 1976. "Practical Reasoning." Reprinted in *The Philosophy of Action*, edited by Alfred Mele, 149–177. Oxford: Oxford University Press, 1997.
Hornsby, Jennifer. 2001. *Simple Mindedness*. Cambridge, MA: Harvard University Press.
Horst, David. 2012. *Absichtliches Handeln*. Paderborn: Mentis Verlag.
Lavin, Douglas. 2013. "Must There Be Basic Action?" *Noûs* 47 (2): 273–301.
Marcus, Eric. 2012. *Rational Causation*. Cambridge, MA: Harvard University Press.
Mayr, Erasmus. 2011. *Understanding Human Agency*. Oxford: Oxford University Press.
McDowell, John. 1994. *Mind and World*. Cambridge, MA: Harvard University Press.
Mele, Alfred. 1992. *Springs of Action*. Oxford: Oxford University Press.
Millikan, Ruth G. 1989. "In Defense of Proper Functions." *Philosophy of Science* 56: 288–302.
Neander, Karen. 1991. "The Teleological Notion of 'Function'." *Australasian Journal of Philosophy* 69: 454–468.
Neta, Ram, and Guy Rohrbaugh. 2004. "Luminosity and the Safety of Knowledge." *Pacific Philosophical Quarterly* 85: 396–406.
Plantinga, Alvin. 1993. *Warrant and Proper Function*. Oxford: Oxford University Press.

Pritchard, Duncan. 2005. *Epistemic Luck*. Oxford: Oxford University Press.
Raz, Joseph. 2011. *From Normativity to Responsibility*. Oxford: Oxford University Press.
Rödl, Sebastian. 2007. *Self-consciousness*. Cambridge, MA: Harvard university Press.
Schlosser, Markus E. 2007. "Basic Deviance Reconsidered." *Analysis* 67 (3): 186–194.
Searle, John. 1983. *Intentionality*. Cambridge: Cambridge University Press.
Searle, John. 2001. *Rationality in Action*. Cambridge: Cambridge University Press.
Setiya, Kieran. 2007. *Reasons without Rationalism*. Princeton: Princeton University Press.
Smith, Michael. 1987. "The Humean Theory of Motivation." *Mind* XCVI: 36–61.
Sosa, Ernest. 1999. "How to Defeat Opposition to Moore." *Philosophical Perspectives* 13: 141–153.
Steward, Helen. 1997. *The Ontology of Mind*. Oxford: Oxford University Press.
Thompson, Michael. 2008. *Life and Action*. Cambridge, MA: Harvard University Press.
Velleman, J. David. 1989. *Practical Reflection*. Princeton: Princeton University Press.
Velleman, J. David. 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press.
Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.