

Agent Causation in a Neo-Aristotelian Metaphysics

JONATHAN D. JACOBS AND
TIMOTHY O'CONNOR

Freedom and moral responsibility have one foot in the practical realm of human affairs and the other in the esoteric realm of fundamental metaphysics—or so we believe. This has been denied, especially in the metaphysics-bashing era occupying the first two-thirds or so of the twentieth century, traces of which linger in the present day. But the reasons for this denial seem to us quite implausible. Certainly, the argument for the general bankruptcy of metaphysics has been soundly discredited. Arguments from Strawson and others that our moral practices are too deeply embedded in human life to rest on anything as tenuous as a metaphysical doctrine far from the thoughts of ordinary people would seem to prove too much: we can easily imagine fantastic scenarios far from the thoughts of ordinary people—involving, say, alien manipulation or massive deception—that, if true, would clearly undermine claims to freedom and responsibility. For still other philosophers, the separation of the moral life from (some) metaphysical issues is prescriptive, not descriptive: it is a recommendation that we revise ordinary moral thought by severing its allegedly problematic links to metaphysics. (Some philosophers appear to hover undecided between such a prescriptive project and a Strawsonian descriptive claim.) We suspect that the prospects of retaining the binding force of ordinary moral thought, were such a reconceived moral practice widely embraced, are bleak. A transition to something closer to moral nihilism seems at least as likely. In any case, our interest here is in descriptive metaphysics, not revisionary.

To say as we do that freedom and moral responsibility have a partly metaphysical character is not to suggest that they can be had only if some highly specific version of a particular metaphysical framework is correct. Instead, we suggest in what follows, it is a *broadly* neo-Humean metaphysics that is not hospitable to freedom (for reasons distinctive to the metaphysics), while a *broadly* neo-Aristotelian metaphysics is. But we also think (and it is the main aim of our paper to show) that different versions of the neo-Aristotelian metaphysics lead to rather different metaphysical accounts of free and responsible action. Specifically, we will argue that (1) the most satisfactory account of human freedom within the broadly neo-Aristotelian metaphysics is agent-causal, but that (2) two different versions of the general metaphysics will lead to important differences in the agent-causal account of freedom. Adjust the details of your general metaphysics, and the details of your account of freedom are transformed in significant ways. Action theory cannot properly be pursued in isolation from general metaphysics.

1. Freedom and neo-Humeanism

David Lewis popularized a certain form of neo-Humean metaphysics, according to which causal facts and the laws of nature are reducible to facts concerning the global spatiotemporal arrangement of fundamental natural properties (which we allegedly may conceive in non-dispositional terms). Roughly, the laws are the best system of generalizations over such natural facts, where bestness is determined by the optimal balance of simplicity and “strength” (explanatory power). Causation in turn consists in a restricted kind of counterfactual dependence of one event on another, where the counterfactuals are grounded in cross-world similarities.¹ There are well-known problems with counterfactual accounts of causation, but we will not render any pessimistic verdict here.² Furthermore, the problem that

¹ The locus classicus is Lewis's article “Causation,” reprinted in 1986. (We note that Lewis allows for temporally remote causation by defining causal chains in terms of stepwise counterfactual dependencies, but it is unnecessary to fuss about such details here.)

² For discussion, see the essays in Collins et al. (2004), which includes ‘Causation as Influence’, in which Lewis proposed a revision of his theory. Hitchcock (2001) and Woodward (2003) have defended rather different counterfactual accounts that employ the structural equations framework that was given a major articulation and development by Pearl (2000). For discussion of these developments, see Menzies (2008).

we do press against a neo-Humean account of free action is not dependent on a counterfactual theory of causation. It is a problem for any reductive account of causation, and we discuss Lewis's picture simply for the sake of concreteness.

Within the neo-Humean framework, intentional agency is naturally understood in terms of the counterfactual dependence of behavior or behavior-guiding intentions on appropriate beliefs, desires, or intentions the agent had immediately before and as the behavior occurs. That human beings act is (nearly!) uncontroversial. That we act *freely* can more plausibly be questioned. We assume here that both metaphysical freedom and moral responsibility are incompatible with causal determinism. Necessary conditions on free actions include plausible compatibilist constraints (e.g., the absence of strong internal or external compulsion) and that they are not determined to occur over some interval terminating in the initiation of the action.

The inclusion of a non-negligible degree of indeterminism in one's account of the proximate genesis of free actions is thought by many to give rise to problems of explanation and control. But questions of explanation and control are better posed within particular metaphysical frameworks. It seems to us that if the neo-Humean framework is accepted, indeterminism need not present a *special* problem of control. Causation is just counterfactual co-variation of a certain kind, and the neo-Humean can readily describe a form of co-variation of motivational factors and behavior that applies to the indeterministic case.³ (Indeed, this fact has been insufficiently recognized by compatibilists who have held that something approximating determinism is necessary for freedom.) We should require only that the objective *chance* of the behavior's occurring would have been much less in the absence of those factors. Furthermore, the counterfactual dependence of the chance of behavior on psychological facts with which the agent identifies is all that it *could be* for a person freely to form a choice. (Irreducible agent causation, for example, makes no sense in this metaphysics, so its omission can hardly be judged a deficiency.⁴) Hence, a suitably textured, causally indeterministic theory of free action gives everything that a neo-Humean could sensibly

³ For an excellent discussion of this issue, see Clarke (1995).

⁴ Obviously, we are further assuming, though less contentiously, that whatever broad metaphysical account of contingent reality is correct for our world will hold for all worlds involving contingent concrete particulars. It is not the case that some worlds are neo-Aristotelian while others are neo-Humean. Without this assumption, the neo-Humean account might well be deficient on grounds that there is a kind of direct control of action had by some *possible* agents though by no agents in a neo-Humean world.

want for an account of metaphysical *freedom* (a fact that is insufficiently recognized by some agent causationists).⁵

We defer consideration of what is or is not *explainable* with respect to undetermined action, treating it in the context of our preferred libertarian account of freedom. It is true that, given a position held by some neo-Humeans (and others as well) that there are special explanatory limits in indeterministic worlds, there will indeed be a serious problem of explanation facing any indeterministic account of agency. But that position is not compulsory, and we will suggest below that it is implausible.

In our view, the above neo-Humean account of free agency is founded on a deeply problematic general thesis of causal reductionism. By taking the fact of A's being a cause of B to be a reducible, massively extrinsic relation—grounded in what occurs elsewhere and elsewhere—we empty the fundamental idea that causes “produce” or “bring about” their effects of any clear content.⁶ Since agency is a causal notion, this problematic consequence carries over: on a neo-Humean analysis, the sense in which my beliefs and desires here and now *bring about* my present action is at best very weak tea. *A fortiori*, extrinsic analyses, on which whether or not psychological factors are causes of behavior is metaphysically determined in large measure by what happens in the distant reaches of spacetime, provide a bizarre account of a *free* action's being, as we commonly say, “directly controlled by” the agent, such that it was “up to her” what she would do in the particular circumstances.⁷ Our ordinary sense of control with respect to freedom of

⁵ Indeed, the extrinsic grounding of particular causal facts in the neo-Humean framework might lead one to doubt the necessity of indeterminism for freedom. See Beebe and Mele (2002). Unfortunately for the neo-Humean, this same extrinsicity renders it doubtful as an account of causation generally and of agency in particular, as we argue immediately below.

⁶ We should acknowledge that “causation” in folk usage probably cannot be neatly lined up with a fundamental relation in the world, on any likely metaphysical account. The folk, for example, often speak of causation by *absences*, as when one says that Susan's failing to water her neighbor's plant caused it to die. On any plausible metaphysical account, there simply are no absences available to stand in a fundamental relation. In our view, it is most plausible to suppose that ordinary causal talk only roughly tracks an important fundamental relation in the world, which, to avoid contentious semantic disputes, we may call “M-causation.” The folk speak truly (often enough) even when speaking of causation by absences. But such truths are grounded in facts concerning “positive” circumstances that stand in the M-causal relation (whose nature we sketch in the next section). Conversation with Gunnar Björnsson has helped clarify our own thinking here. There is also a nice discussion of this matter in Ted Sider's *Writing the Book of the World* (2011), 15–16 and 75–6.

⁷ See O'Connor (2009) for a development of this point. Gunnar Björnsson has pointed out in discussion that it is open to the neo-Humean to modify her account as follows: our concept of natural law require there to be some minimal score on the balance of simplicity and strength. In neo-Humean

action manifestly points to something that supervenes on the local circumstances in which we act—or, at any rate, circumstances much more local than those thousands of years in the past or future.

2. A Neo-Aristotelian Metaphysics and Event-Causal Libertarianism

There is more than one path away from the neo-Humean's causal reductionism. Here we will consider only the path that we favor: a neo-Aristotelian metaphysics that assigns a central role to primitive causal powers. On this view, natural properties are, or of necessity confer, causal powers on their bearers.⁸ While the neo-Humean's properties are intrinsically inert, the neo-Aristotelian's are intrinsically powerful.⁹ If the neo-Humean world is ungoverned, since laws are merely descriptions of contingent regularities, the neo-Aristotelian world is self-governed, since laws are necessary descriptions of the powerful natures of properties.¹⁰ Neo-Humean causation is a sort of counterfactual co-variance, but neo-Aristotelian causation is the exercise of an irreducible causal power.¹¹ The details of the broadly neo-Aristotelian metaphysics need not concern us here. What are important are the general ideas, first, that properties are (or confer) primitive causal powers and, second, that causation is the exercise of such powers.

In order to understand the nature of *indeterministic* causation within the neo-Aristotelian metaphysics, it is helpful to contrast it with another sort of picture that some contemporary philosophers endorse. On the latter, causal

worlds where the patterns in one cosmic neighborhood or compact world-segment (such as the one we currently occupy) sharply differ from those in others, we should say that the laws themselves vary from one world segment to another. In this way, we needn't say that what contingently occurs in very remote regions of spacetime are needed to fix what actions I bring about (or whether I ever so much as act at all).

We grant that reducing the extent of extrinsicity serves to improve the view. But since it is the very nature of the view to give an extrinsic account of causation, this move cannot make the implausibility go away. There are ever so many neo-Humean worlds where memories and seeming historical traces are radically misleading beyond a short threshold into the past and where the patterns will abruptly change or simply cease in the very short future. Whether or not these things are in fact so just seems beside the point when we ask whether a present bodily motion is something that I freely bring about. (Note that the point concerns metaphysical determination, not epistemic justification.)

⁸ See, e.g., Shoemaker (1980, 1998), Heil (2003), Mumford (2004), Bird (2007), and Martin (2008).

⁹ For full discussion, see Jacobs (2011).

¹⁰ See, e.g., Mumford (2004) and Bird (2007).

¹¹ See, e.g., Martin (2008), Mumford (2009), and Bird (2010) for discussion.

indeterminism is thought of as *causation of probability*. Indeterministic causes, no less than deterministic ones, are always and uniformly efficacious. They do not cause the undetermined *outcome* of an indeterministic process, but instead cause the outcome's objective chance of occurring (generally by raising the probability, though in certain kinds of cases the probability may actually be lowered). Beyond helping to fix the prior chances of an event, there is nothing more that a cause does. Where the chance is 1, the cause suffices for the effect, and so we naturally, if misleadingly, say that it brings about the effect. But strictly speaking, nothing brings about the effect, whether the chance is 1 or less than 1. Only the prior chance is brought about.

This causation of probability view is perhaps assumed (consciously or not) in objections to the intelligibility of indeterministic agency. If an outcome is not brought about by anything, it's hard to see how it can be something that the agent controls and that we may fully explain in terms of her reasons for acting. But we should reject the *causation of probability* interpretation in favor of a *probability of causation* alternative precisely because the former makes the occurrences of events in indeterministic worlds utterly mysterious. There is no reason within a causal powers metaphysics to suppose that causes must always produce their characteristic effects, so that in indeterministic scenarios we have to resort to the fiction of regular causings of objective chances. We should suppose instead that indeterministic causes produce their effects though they need not have done so: they are *propensities* towards a plurality of possible effects. They are sufficient for each of them only in the sense that they are all that is needed, not in the sense that they are a causally sufficient condition.¹² Every indeterministic event is produced, though none is necessitated. Causation, whether deterministic or indeterministic, is a singular relation—the very same relation. The prior probability of one event's causing another (with limit case of 1) is simply a measure of the strength of its (single-case) propensity to do so, which helps to fix applicable laws of nature.

Let us apply this understanding to an indeterministic account of human free action that, like the neo-Humean account above, is rooted in a causal theory of action generally. According to it, when an agent freely acts, her web of motivational states is jointly disposed towards two or more choices,

¹² Anscombe (1971) famously develops this point.

to varying degrees. Whichever choice is made, it will have been caused by some relevant motivation of the agent, a motivation with which she identifies.¹³ The exercise of agent-control consists in the causal efficacy of one's motivations, and freedom further requires the openness of the future to (or consistency of the past and laws with) a plurality of specific outcomes. Such is a plain vanilla version of event-causal libertarianism.¹⁴

While agents, on this account, do not have any less control over what they do than agents in a corresponding deterministic scenario, they also do not have more. Indeterminism in the causal link between motivations and choice opens up a plurality of alternatives unavailable on determinism, but the agent does not seem to *settle* which of the options is taken in a sense robust enough for the agent to be morally responsible. Autonomous control seems to require more than compatibilist control plus plural alternatives. Consider two event-causal libertarian universes, whose histories have been precisely the same until a time at which two intrinsically identical agents (including psychological propensities towards the same possible choices with the same degrees of strength) make diverging choices. It does not seem correct to say that it was up to the respective agents, something that they were individually responsible for, that one chose the path of insult and the other that of gracious forbearance. It's not that choices in these worlds would be "freakish," the "result of pure chance," and so not something that the agents in any sense *did*. It's merely that the control that is exercised is of an insufficient variety to ground robust freedom and responsibility.

But this objection presupposes the intelligibility of a stronger, more robust variety of control. Unlike on the neo-Humean metaphysics, there *does* seem to be space on the neo-Aristotelian account for such an alternative, as we will now show.

¹³ The condition that the causing motivation be one with which the agent "identifies" is intended to handle possible cases where an agent might be subject to a powerful and perhaps momentary "alien" desire. We needn't concern ourselves here with different accounts of this notion of "identification."

¹⁴ The foremost recent defender of this theory, Kane (1996), augments the account with further conditions on the process by which reasons result in choices. Ekstrom (2000) locates the requisite indeterminism in a special subset of actions—those in which an agent critically evaluates her own conception of the good and comes thereby to have certain preferences that regulate ordinary actions. These proposals are interesting and it is worth considering the issues that they raise in their own rights. But they do not, in our judgment, suffice to answer the fundamental concern with event-causal libertarianism that we raise immediately below.

3. The Standard Agent-Causal Alternative

Agent causalists maintain that freedom requires a distinct, enhanced *kind* of control from the causal efficacy of internal states with which one identifies. Responsibility-grounding control resides in an indeterministic, ontologically fundamental causation of a choice or action-guiding intention by the agent. Taking a feature as a metaphysical primitive is a reliable way to ensure that one's overall theory really does allow for the feature, instead of offering a pale substitute in the manner of various implausible reductionisms. And the longstanding difficulty of giving a plausible analysis of our pre-theoretical notion of autonomous control suggests that the gambit of primitive posit is not simply absurd.

We should be careful to distinguish the agent causationist's position from that of non-causalists (e.g., Ginet 1990, Goetz 1988, McCann 1998, and Pink 2004). Both positions agree that autonomous control rests on a primitive capacity to form intentions (or volitions, according to the theorist's preference). But the agent causationist insists that this capacity is—and can only be—causal in nature. The non-causalist, by contrast, ascribes 'active power' or 'the power of choice' to the agent while insisting that these terms are to be understood non-causally. However, it is unclear to us what this means. It seems to us that the term "power" is being misappropriated for rhetorical purposes. Better that these theorists simply say that nothing causes free choices or volitions but that, notwithstanding, which choice is made is controlled by the agent, in virtue of the fact that the choice is his.¹⁵ Such a statement is clearer—though clearly false, in our estimation.

In assessing the agent-causal account, we need to consider the role of the agent's motivational states in the production of their undetermined choices. Randolph Clarke (2003) proposes an "integrationist" account on which free actions are caused *both* by the agent (*qua* substance cause) and by certain of the agent's motivational states (*qua* event cause). Clarke proposes that, in the presence of a "live" agent-causal capacity, it is a law of nature that:

¹⁵ It is worth noting that defenders of libertarian accounts of freedom that are ostensibly event causal sometimes respond to the problem of control by emphasizing not that the choice is caused by the agent's reasons but simply that it is "his"—it occurs within the agent. We think that the tendency of event causalists when pressed to shift between a causal and a non-causal, "ownership" account of control is revealing.

- (a) whatever action is performed will be caused by the agent,
- (b) a particular reason will cause an action only if the agent causes it, and
- (c) the agent will cause an action only if some corresponding reason also causes it.

It seems to us that, absent further explanation, a lawful and symmetrical causal yoking of *this* sort is mysterious. Surely one or the other causal factor will be in the driver's seat (given, as Clarke says, that they are not each partial causes). And we want there to be at least one sort of explanatory asymmetry: it is because the agent had those reasons that he (*qua* agent cause) caused the action that he did, not the other way around. Yet Clarke can't say that the state of having those reasons indeterministically *brought about* the agent-causal event on pain of making indeterministic causation by reasons more fundamental than (because prior to) agent causation in the production of an action. The resulting account would seem to offer no improvement over a simple event-causal account that dispenses with primitive agent causation.¹⁶

O'Connor (2008) suggests a different account of the way that reasons influence agent-causal actions. He suggests that while agent-causal events are *unproduced* by other events, they are probabilistically *structured* by myriad factors, especially the agent's own motivational states. As a result, agents have a continuously evolving, objective propensity to cause intentions to act in ways they take to be suited to their ends. More carefully, the idea is that motivational states act causally on the persisting capacity of an agent freely to form an intention to act, altering the objective strength of (or generating) the dispositions the agent has to form specific intentions within certain intervals. The influence of reasons so conceived is not unlike how things go according to the *causation of probability* interpretation of indeterministic causation. However, O'Connor's account of the influence of reasons on agent-causal choices is not offered as an account of the nature of causation itself, and it does not have the absurd consequence that nothing brings about the specific outcome of an indeterministic process. This account has the advantage (over Clarke's integrationist account) of offering a unified picture of the flow of causal influence, and it does so without sacrificing the core agent-causal commitment to the exercise of a power that is not itself in turn produced by previous events.

¹⁶ For further discussion, see O'Connor and Churchill (2006).

Note that on this neo-Aristotelian framework, having reasons (understood as motivational states) is having certain kinds of causal powers.¹⁷ So, as an agent first comes to have reasons for a course of action that was previously (subjectively) unmotivated, she comes to have new powers of choice and action. In a case where she comes to have additional reasons for an already motivated action type, the power so to act is not altered. However, insofar as the different reasons also motivate different actions of however fine-grained a type, new powers are thereby acquired. And distinct reasons must have the potential to motivate somewhat different action types under at least some possible circumstances, for otherwise they could not be individuated within the causal powers framework.

On O'Connor's account of agent-causal power, there is one persistent agent-causal power, a power to form an intention to act. And an agent with that power can have differing specific propensities so to act, depending on what reasons the agent has. The reasons are, in part, powers to act on the persistent agent-causal power, to alter its strength. It is worth noting, however, that the causal powers metaphysics does not, by itself, entail O'Connor's view of agent-causal power. Indeed, on some versions of the causal powers metaphysics, specific propensities are essential to causal powers.¹⁸ On this way of thinking about it, when a substance has a power to do some action, A, the power is a power to do A with some specific probability in specific circumstances. Were the probabilities different, even slightly, it would be a different power. Hence, on this view of powers, reasons are powers to bring about *various* agent-causal powers, each conferring specific probabilistic tendencies towards specific outcomes in specific circumstances. Speaking loosely, one has "agent-causal power." In strict truth, however, there are a family of related agent-causal powers; all such powers are similar in being powers of the agent to bring about an intention with some specific content. This issue, however, is orthogonal to our central topic, as either view—O'Connor's view of a persistent agent-causal

¹⁷ "Reasons" can refer to *normative reasons*, or the conditions (generally external to the agent's psychological states) that rationally or morally justify a particular course of action for an agent in a given circumstance, whether or not the course of action is taken or the agent even acknowledges the existence of the reason. "Reasons" can also refer to *motivational reasons*, the agent's own reasons for doing what he does, wise or foolish as may be. In this latter sense, having a reason is a psychological state or set of states (such as beliefs, desires, and intentions) that motivates the agent towards and potentially explains certain courses of action. It is this latter, motivational sense of "reasons" that is in view here.

¹⁸ See, e.g., Jacobs (2011).

power, or the alternative view of a family of related agent-causal powers—can be accepted in both the above causal powers metaphysics and the modified causal powers metaphysics to which we turn now.

4. A Modified Causal Powers Metaphysics

On the above causal powers metaphysics, the causes of events that do not involve agents, and indeed of many events that do involve agents, are events. The addition of agent causation to such a picture therefore involves the addition of a *new kind* of causation. But there is an alternative account of causation that fits well within a causal powers metaphysics, on which the causes of all events are *substances*. All causation, on this view, is substance causation. E. J. Lowe (2008) argues for this view roughly in the following way. Causation is the exercise or manifestation of a power. The cause is the thing that has the power. But only substances have powers. Therefore, only substances are causes. Events are the having of a power by substances, and those powers are exercised or manifested by substances.

On Lowe's version of the substance causation view, substances cause effects by manifesting a power. But whenever a substance causes an effect, it does so ultimately by manifesting a *non-causal* power. The rock caused the tree's breaking, by rolling into it. And the rolling of the rock is a manifestation of the rock's non-causal power to roll. In other words, the rolling of the rock does not consist in the rock's causing anything. It consists in, say, the rock's changing position.

In many cases, such manifestations of non-causal powers may themselves be effects. Not so with what Lowe calls spontaneous powers: when a substance manifests its spontaneous power, it is not caused to do so by anything, and its manifestation of the spontaneous power does not consist in its causing anything. Such is Lowe's account of both radioactive decay and free action. In both cases, the fundamental source of action is the exercise of a non-causal, spontaneous power. That is to say, the fundamental action, by which the agent or atom *does* cause something, when it does, is an event that is neither a causing nor caused. In the case of free action, it is a willing. A willing is not a causing—that is, it does not consist in the substance causing an effect. And it is not caused by anything. Still, the agent does cause something, say, raising her hand, by willing to raise her hand, when her willing is effective.

In our judgment, Lowe's analysis unhelpfully complicates the substance-causal powers metaphysics. By introducing a fundamental distinction between the manifesting of a power and that by which a substance manifests the power, he invites a question concerning the nature of the "by" relation that has no satisfactory answer that we can see. And it is unclear how this posit offers any improvement over a simpler analysis on which a substance's causing an effect simply *is* its exercising a causal power. What's more, Lowe's view leads to an even clearer problem in the account of freedom. For central to the account is a non-causal sort of power, something that we encountered above in discussing non-causalism. Lowe's uncaused volitions appear to be no different, intrinsically, from the non-causalist's volitions or choices. As there, so too here: it just seems misleading to call the spontaneous occurrence of such events "the exercise of non-causal *powers*," given that the events have no causes and no internal causal structure. We think it doubtful that control can be understood in non-causal terms.

For these reasons, we find the following analysis to be preferable: a substance's having a property is its having a causal power of a specific sort. A substance's causing an effect is its manifesting such a power or its co-manifesting a power with other substances. (Note that on this view, causation is non-transitive, since causes are substances and effects are events.) In some possible cases, given the totality of properties had by an object and its situation, the effect is causally determined to occur. The conjunction of interacting powers yields a probability of 1 that the substance or substances will cause that very type of effect. In other possible cases, the effect is causally undetermined. Here, the exercise of more than one power is possible and presumably each is probable to some specific degree.

The modified neo-Aristotelian ontology, then, is one on which substances have powers, and all and only substances are causes of effects. When a substance causes an effect, it exercises its power to do so—that is, a substance's causing of an effect is identical to its exercising its power to bring the effect about.

5. Agent-Causal Libertarianism and the New Ontology

Some of the basic claims of the agent causationist concerning freedom carry over from the event-causal to the substance-causal powers metaphysics. First, agents can be literal causes of actions. (Better: actions are agent-causings

of intentions.) Second, nothing produces an agent's causing of an intention. (Since on this metaphysics, in general, nothing produces any causing of an event by a substance, whether its activity is determined or not.) Finally, the motivation for incompatibilism about freedom, such as it is, remains.¹⁹

However, the substance-causal powers metaphysics forces some changes to an agent-causal account of freedom. On the event-causal powers metaphysics, reasons either structure the agent-causal power, in O'Connor's sense, or they cause the agent to have the specific agent-causal power she has. What do reasons do, according to the modified framework now under discussion? Strictly speaking, nothing. The agent (and the particles that compose her—see below) do things. Some of those doings are the exercise of a fundamental agent-causal power to form intentions to act which the agent has in part because she has the reasons she does. So, reasons are causally *relevant*. She wouldn't have done what she did, and wouldn't have been capable of doing what she did, were she not to have those (or other) reasons so to act. But, strictly speaking, reasons are not causes, since reasons are not substances, and only substances are causes.

As noted in the previous section, O'Connor (2000, 2008) proposes from within an event-causal powers framework that what reasons do is to confer "carried propensities" or "tendencies" on a generic and persisting agent-causal capacity. This causation of probability has to be taken on board as a kind of influence that differs from the probability of causation that characterizes non-intentional indeterministic causes. An advantage of the framework now under consideration is thus ideological simplification on this point.

We also have the further ideological simplification resulting from the fact that agent causation is not a fundamentally distinct kind of causation. What is distinctive about agent causation among other varieties of substance causation on this view is merely that the cause is conscious, intentional, and freely chooses the ends for which it will act.²⁰ As we see it, each of these

¹⁹ We note the interesting fact that the substance-causal metaphysics enables one to make good sense of Markosian's (1999) claim that agent causation is consistent with compatibilism. (Whereas it is not clear that it is compatible with the more common event-causal powers metaphysics.) We disagree with Markosian, however, that compatibilism becomes more attractive once one endorses an agent-causal theory of free action. Once conceptual space is opened up for determined agent causation, the question of whether there is a substantive distinction between being produced by me and being *freely* produced by me is on the table. As we see it, a suitably formulated version of the Consequence Argument for incompatibilism is compelling (see O'Connor 2000, Ch. 1).

²⁰ Ruth Groff suggested a similar view in conversation.

distinctive sorts of capacity is ontologically fundamental. (They are interconnected in certain ways, with the last, in particular, presupposing the first two.) “Fundamental” here means that having specific instances of these types of capacity does not consist in (and is not “constituted” or “realized” by) the agent’s having some set of other capacities, or in the agent’s parts’ having certain capacities and standing in certain relations. Given (what we take to be) the fact that human agents are composed systems, we must think of human persons as ontologically emergent substances. Among the powers of our fundamental parts are powers collectively to cause system-level properties/powers, powers that are sustained as long as the system (the person) retains the requisite form of organized complexity. As the bearers of fundamental powers, the person is an ontologically fundamental, albeit composed substance.²¹ Agents *freely* act because they (literally) cause their effects with the conscious aim of attaining certain ends, and their doing so is not settled by features of their situation (whether external or internal) up to the time of the action.

We have, then, two versions of agent causalism, one embedded with an event-causal, neo-Aristotelian metaphysics, the other within a substance-causal, neo-Aristotelian metaphysics. On the first view, agent-causal power is a power to cause things in a fundamentally distinct sort of way, involving agent causation rather than the typical event causation. The reasons that an agent has are powers to structure her agent-causal power, either by causally influencing the strength of a persisting agent-causal power or by bringing about a new, slightly different agent-causal power. On the second view, when an agent causes an event, it does not involve a unique sort of causation. All causation is substance causation. Rather, the uniqueness comes from the sort of substance, a conscious substance influenced by reasons, and the sort of effect, an intention to act for a certain reason.

6. Revisiting the Alleged Problems of Control and Explanation

Some philosophers contend that agent causation, even if coherent, cannot solve the causal indeterminist’s problems of explanation and control. We

²¹ See O’Connor and Jacobs (2003) for a detailed account of this picture of ontologically emergent substances within an event-causal powers metaphysics.

consider a typical way that each of these problems is presented against agent causation and argue that they fail.

Agent Causation and the Problem of Explanation

Consider first the problem of explanation. Where an event is undetermined, there was some objective chance that an alternative type of event might have occurred—the alternative’s obtaining had a non-zero causal probability in the total set of circumstances. This implies that any cause or causes that one might cite in a putative explanation of the actual event was consistent with the occurrence of the alternative. It appears to follow that the cause cannot explain why the actual event obtained *rather than* the possible alternative. From this it might seem to follow that the cause cannot, after all, ‘fully’ explain the actually occurring event itself, since to fully explain why an event occurred is *inter alia* to explain why it occurred rather than any alternative.

Against this argument, we note that not all causal explanations of events must be contrastive or imply the availability of contrastive explanations, for every possible contrast. As Peter Lipton (1990) made clear, a request for a contrastive explanation (“Why P rather than Q?”) presumes that there is an explanatory relationship between fact (P) and “foil” (not-Q); it presumes that the occurrence of P and the non-occurrence of Q can be given a unifying explanation. But this assumption plainly will not hold for every such pairing even in a deterministic world—as when the occurrence of P and the absence of Q are completely unrelated matters. In an indeterministic world, contrastive explanation will also fail (plausibly) wherever P and Q are mutually exclusive; each had a substantial chance of occurring, and P was not significantly more probable than Q. *But it does not follow that there can be no explanation of P, or that whatever non-contrastive explanation there may be of P will be somehow deficient—of a lesser variety of explanation than contrastive explanation.* We explain—really explain—an indeterministic outcome P by citing and describing the causal factor or factors that brought it about, including the cause’s objective set of objective propensities and the most salient, proximate causes of its having such propensities.

The point is a familiar one in scientific explanations of indeterministic phenomena unrelated to free action. If there are a plurality of possible outcomes of the interaction of a pair of particles, the particular outcome

that obtains has an explanation in terms of propensities of the two particles which actually were manifested, bringing about that particular result. Once one understands the indeterministic nature of those propensities and others that were not, but might have been, manifested on that occasion, one realizes that there is nothing further to explain about the situation.

We can even explain why there can be no true contrastive explanation of the fact that P occurred rather than Q by underscoring the indeterministic nature of the causal source in question. Philosophers in the grip of the Principle of Sufficient Reason profess mystification at this scenario, but don't give any kind of argument. If we grant that there can be indeterministic causal mechanisms (or agents), then deterministic causes are just the limit case of a continuum of probabilistic causes, and which sorts of explanation it is appropriate to seek depends on which sort of world we occupy.

The application of this general point to our account of human free action is as follows: as we come through various causes to have motivations to act in various ways, the interplay of these motivations and other influences result in an array of propensities to choose and act of varying strengths. Suppose that while deliberating on what to do on a Saturday afternoon, I am disposed with a strength of 0.3 to help a friend repair her deck, owing to my awareness that she wants to get it done soon, could do so more easily if she is helped, has helped me in similar ways in the past, together with my desires to be and to be perceived as helpful to her. And suppose that I am disposed with a strength of 0.7 to watch a football game instead, for the obvious reasons. I choose to help my friend. Question: Why did I so choose? Answer: I so chose because I was motivated by my awareness that my friend wanted to get the job done, etc. That is to say, those beliefs and desires were the predominant factors determining my propensity so to choose. And that is a perfectly good explanation of why I chose as I did, even though there will not be an explanation of why I chose to help rather than to watch the football game.²²

²² It is worth making explicit here how we would respond to Davidson's (1963) famous challenge to theories of (free) action according to which reasons are not causes. (As we noted earlier in the text, on our view, motivational reasons are not causes, strictly speaking.) Davidson asks how, if reasons are not causes, we will distinguish cases in which an agent has distinct reasons A and B for choice C but does so for reason A from similar cases in which she does so for B, and from still other cases in which she does so for both A and B. But our view supplies a ready answer: the agent acts for all and only those reasons she had which made a non-negligible contribution to her propensity to choose as she did.

Agent Causation and the Problem of Control

Recall the argument directed against the causal indeterminist that was based on a comparison of indeterministic worlds identical up to a certain time at which intrinsically identical agents make diverging choices. As the causal indeterminist conceives things, agents don't cause events—only events do. Where choices are causally undetermined, some motivational state of the agent brings about the choice though it was possible that another state have brought about a different choice. Hence, one set of motivations cause the choice to insult in world W_1 , while a distinct set of motivations cause the choice to refrain in W_2 . We judged that in this scenario, it is not up to the respective agents, something that they were individually responsible for, to make the choices they do. Indeterminism of this sort confers opportunity but not an enhanced form of control that enables the agent responsibly to exploit the opportunity.

Some argue that the agent causationist can do no better (Haji 2004; see also the related, intra-world “rollback” argument in van Inwagen 2000). For here, too, there is *nothing* whatsoever *about* the one agent, right up to the moment of the choice, that distinguished her from the other, and so nothing about her that made the difference to what she did. Each had the same propensity to insult or refrain. Once again, it seems that neither agent controlled the way their respective cases unfolded in such a way that it was up to her that she spoke the insult or refrained.

How one replies to this objection depends, in part, on whether one accepts the event-causal or substance-causal powers metaphysics. On the former, there is available the very simple reply that the objection fails to take seriously the concept of agent causation, which is conceived on this metaphysics as a primitive form of control over undetermined, single-case outcomes. The agent's control is exercised not through the efficacy of *prior* states of the agent (as on causal theories of action), but *in* the action itself. Susan's causing her intention to publicly insult her opponent is itself an *exercise* of control. And since, *ex hypothesi*, it is quite literally the agent herself generating the outcome, it is hard to see how the posited form of control could possibly be improved upon.

But on the substance-causal powers metaphysics, agent causation is not a special sort of causation, and so cannot be a special, agential form of control simply by dint of the basic sort of causation manifested. Return, then, to the

two indeterministic worlds, W_1 and W_2 , identical up to a certain time at which intrinsically identical agents make diverging choices. And let us consider three such pairings: a neo-Humean pair of 'worlds,' an event-causal libertarian pair of 'worlds,' and a purely substance-causal pair of 'worlds.' (We put 'worlds' in single quotes to mark the fact that it will be disputed whether any of these descriptions match genuinely possible worlds.) As on the event-causal powers metaphysics, the proponent of the substance-causal powers metaphysics can insist on the significance of the fact that only in the latter pair of worlds are agents literally the causes of their choices. This fact, construed narrowly, is not *sufficient* to ground the claim that agents in these substance-causal worlds but not the others are responsible for their choices, since it is possible (and indeed common) for substances to cause their effects without purpose, consciously or otherwise. But it does mark a relevant and important difference that, when combined with the fact that the causation is consciously goal-directed, grounds a more robust variety of control than is possible in either neo-Humean or purely event-causal neo-Aristotelian worlds. The agent, herself, consciously and intentionally brought about the effect in both worlds. That, we submit, is a sufficiently robust form of control to ground freedom and responsibility.

7. Conclusion

We have not here endorsed, let alone argued for, the substance causation metaphysics. Our aim was only to explore how embracing it would alter the way that we conceive of metaphysical freedom. There is a more general moral that we hope will become more widely embraced in action theory and philosophy of mind: metaphysics matters.²³ Debates over reductionist, epiphenomenalist, or emergentist alternatives concerning both intentionality and consciousness, and over freedom and determinism, turn more on general metaphysical positions than is commonly acknowledged.²⁴

²³ We are not alone in drawing that conclusion. See Beebe and Mele (2002) and the evolving debate between Jaegwon Kim and his recent critics over non-reductive physicalist accounts of mental causation.

²⁴ Jonathan Jacobs worked on this project during an NEH summer seminar, "Metaphysics and Mind," led by John Heil in 2009. He would like to thank the NEH for its support, and John Heil, Jonathan Lowe, and the participants in the seminar for their feedback. He would also like to thank the John Templeton Foundation, for a grant supporting work on this project.

References

- Anscombe, E. (1971). *Causality and Determination: An Inaugural Lecture*. Cambridge: Cambridge University Press.
- Beebe, H. and A. R. Mele (2002). 'Humean Compatibilism'. *Mind*, 111(442): 201–23.
- Bird, A. (2007). *Nature's Metaphysics: Laws and Properties*. Oxford: Oxford University Press.
- (2010). 'Causation and the Manifestation of Powers'. In A. Marmodoro (ed.), *Powers: Their Grounding and their Manifestations*. Abingdon: Routledge: 160–8.
- Clarke, R. (1995). 'Indeterminism and Control'. *American Philosophical Quarterly*, 32: 125–38.
- (2003). *Libertarian Accounts of Free Will*. New York: Oxford University Press.
- Collins, J., N. Hall, and L. Paul (eds.) (2004). *Causation and Counterfactuals*. Cambridge, MA: MIT Press.
- Davidson, D. (1963). 'Actions, Reasons, and Causes'. *Journal of Philosophy*, 60: 685–700.
- Ekstrom, L. W. (2000). *Free Will: A Philosophical Study*. Boulder, CO: Westview Press.
- Ginet, C. (1990). *On Action*. Cambridge: Cambridge University Press.
- Goetz, S. (1988). 'A Noncausal Theory of Agency'. *Philosophy and Phenomenological Research*, 49: 303–16.
- Haji, I. (2004). 'Active Control, Agent-Causation, and Free Action'. *Philosophical Explorations*, 7(2): 131–48.
- Heil, J. (2003). *From an Ontological Point of View*. Oxford: Oxford University Press.
- Hitchcock, C. (2001). 'The Intransitivity of Causation Revealed in Equations and Graphs'. *Journal of Philosophy*, 98: 273–99.
- Jacobs, J. D. (2011). 'Powerful Qualities, Not Pure Powers'. *The Monist*, 94(1), 81–102.
- Kane, R. (1996). *The Significance of Free Will*. New York: Oxford University Press.
- Lewis, D. (1986). 'Causation'. In *Philosophical Papers: Volume II*. Oxford: Oxford University Press: 159–213.
- Lipton, P. 1990. 'Contrastive Explanation'. In D. Knowles (ed.), *Explanation and its Limits*. Cambridge: Cambridge University Press: 247–66.
- Lowe, E. J. (2008). *Personal Agency: The Metaphysics of Mind and Action*. Oxford: Oxford University Press.
- McCann, H. (1998). *The Works of Agency*. Ithaca: Cornell University Press.
- Markosian, N. (1999). 'A Compatibilist Version of the Theory of Agent Causation'. *Pacific Philosophical Quarterly*, 80(3): 257–77.
- Martin, C. B. (2008). *The Mind in Nature*. Oxford: Oxford University Press.

- Menzies, P. (2008). 'Counterfactual Theories of Causation'. *The Stanford Encyclopedia of Philosophy*. Available at <http://plato.stanford.edu/entries/causation-counterfactual/>.
- Mumford, S. (2004). *Laws in Nature*. London: Routledge.
- (2009). 'Passing Powers Around'. *The Monist*, 92(1): 94–111.
- O'Connor, T. (2000). *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.
- (2008). 'Agent-Causal Power'. In T. Handfield (ed.), *Dispositions and Causes*. Oxford: Clarendon Press: 189–214.
- (2009). 'The Significance of Persons'. In J. Aguilar and A. A. Buckareff (eds.), *Philosophy of Action: 5 Questions*. Birkerød, Denmark: Automatic Press/VIP.
- O'Connor, T. and J. R. Churchill (2006). 'Reasons Explanation and Agent Control: In Search of an Integrated Account'. *Philosophical Topics*, 32: 241–54.
- O'Connor, T. and J. D. Jacobs (2003). 'Emergent Individuals'. *Philosophical Quarterly*, 53: 540–55.
- Pearl, J. (2000). *Causality*. Cambridge: Cambridge University Press.
- Pink, T. (2004). *Free Will: A Very Short Introduction*. Oxford: Oxford University Press.
- Shoemaker, S. (1980). 'Causality and Properties'. In P. van Inwagen (ed.), *Time and Cause: Essays Presented to Richard Taylor*. Dordrecht: D. Reidel: 109–36.
- (1998). 'Causal and Metaphysical Necessity'. *Pacific Philosophical Quarterly*, 79(1): 59–77.
- Sider, T. (2011). *Writing the Book of the World*. Oxford: Clarendon Press.
- van Inwagen, P. (2000). 'Free Will Remains a Mystery'. *Philosophical Perspectives*, 14: 1–20.
- Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.