

Agent-Causal Theories of Freedom

Timothy O'Connor

[For Oxford Handbook On Free Will, 2nd ed., R. Kane (ed.) Draft Jan.25, 2010]

This essay will canvass recent philosophical discussion of accounts of human (free) agency that deploy a notion of *agent causation*. Historically, many accounts have only hinted at the nature of agent causation by way of contrast with the causality exhibited by impersonal physical systems. Likewise, the numerous criticisms of agent causal theories have tended to be highly general, often amounting to no more than the bare assertion that the idea of agent causation is obscure or mysterious. But in the past decade, detailed accounts of agent causation have been offered (chiefly by Randolph Clarke and Timothy O'Connor), and they have occasioned more specific objections in turn.¹ These recent accounts and objections to them will be my primary focus in what follows. But first I will identify two distinct motivations that have been advanced for adopting an agent causal approach to human agency and the ontological and metaphysical commitments common to any version of this approach.

I Motivations for an Agent Causal Account

From the Intractable Difficulties with Giving a Causal Theory of Action

Many action theorists have sought informative necessary and sufficient conditions for a behavior's being an intentional action—something the agent did purposely, not accidentally or reflexively. A common strategy starts from the assumption that psychological states such as desires, beliefs, and perhaps intentions are important and

salient antecedent causes of action. This strategy was championed by Donald Davidson (1963) and has been the dominant approach ever since. Nevertheless, soon after Davidson's essay, a number of authors (Chisholm 1966, Taylor 1966, and Davidson 1980 himself) noticed a serious obstacle to attempts to provide a plausible causal theory. It is easy to conjure up scenarios where one's motivational reasons cause one to perform an action suited to the reasons despite one's not having acted intentionally. Here is Davidson's example:

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never *chose* to loosen his hold, nor did he do it intentionally. (1980, 79)

Here, the climber has a reason for loosening his hold on the rope, and the reason causes him to do so, but in such a way that it is evident that he did not intentionally so act. The *way* the reason causes the action was of the wrong sort for the action to have been intentional. The challenge for the causal theorist, then, is to say in general terms what the *right* way consists in. What kinds of causal process between motivating reasons and behavior must occur for the action to be intentional, according to the causal theorist?

Some philosophers have held that there can be no good answer to this question, so that the "problem of wayward or deviant causal chains" provides a decisive refutation of the causal theory of action. Two different morals have been drawn from this conclusion. According to some, the failure of this approach shows the futility of analyzing intentional

action in fundamentally causal terms of any kind. Instead, actions must be understood in purely teleological terms, through the identification of the goal or purpose for which the behavior was undertaken (Wilson 1989, Sehon 2005). But others have argued that teleological explanations of an action do not provide a plausible analysis of agency itself, and instead require that causation *by the agent* be taken as an ontological primitive in the theory of intentional action (Taylor 1966).

Whether the problem of deviant causal chains is in fact a decisive barrier to giving a causal theory of action is highly contested. Sophisticated attempts to overcome the problem by advancing complicated conditions on the kind of reason-action causation involved in intentional action may be found in Bishop (1989) and Mele (1992). These are exercises in conceptual analysis. See Davis (forthcoming) for the suggestion that an account of the required causal process should come, instead, from a mature psychology and neuroscience. And see Stout (forthcoming) for an argument that the problem is not distinctive to the causal theory of *action*, as there can be irregular links within *any* kind of causal process (such as perception or electrical conduction), which suggests a difficulty for attempts to give a precise and fully accurate causal analysis of the original type of process. We would not abandon an otherwise promising causal theory of these processes in the face of the challenge, so why should we do so in the case of a causal theory of action?

From the Conflict Between a Causal Theory and Freedom of Action

People act intentionally throughout their waking lives. Whether or not they do so freely—in such a way as to render their actions appropriate candidates for moral appraisal—is a

further substantive matter, one that depends on the nature of their control over their own actions. The concept of action is distinct from that of free action, and it is not obvious that a good way to understand freely performed action is to develop a set of plausibly sufficient conditions for action and then to add a further freedom condition. For perhaps there are several interestingly different ways that the concept of intentional action might be satisfied, but some of these do not admit of freedom variants.

Most agent causation theorists are best interpreted in this manner. Agent causation is a necessary feature of freely chosen activity, even though there may be possible forms of intentional activity that lack it altogether. (C.A. Campbell 1967, John Thorp 1980, Alan Donagan 1987, Randolph Clarke 1993, 1996, and Timothy O'Connor 1993, 1995, 2000 explicitly take this view, while Chisholm vacillates in his early essays. Taylor, as already noted, propounded agent causation as a feature of all intentional action, as does Godfrey Vesey 1987, William Rowe 1991, and Richard Swinburne 1997.) Indeed, it may be that while some of our actions are agent-causal in character, others (including habitual and compulsive behaviors) are not. For the remainder of this essay, I will focus solely on this freedom-based motivation for developing an agent-causal account of agency.

Agent causationists have generally also been incompatibilists, holding that freedom of action and causal determinism are incompatible. (Markosian 1999 is an exception, arguing that an agent-causal theory of action enables one to overcome the significant challenges to compatibilism.)

Agent causationists typically hold that the absence of causal determinism in the causal link between prior conditions such as motivational states and the ensuing action is not enough for freedom, or at least for the sort of freedom that can directly ground

ascriptions of responsibility. And what this merely negatively described causal indeterminist scenario lacks is precisely the agent's directly controlling the outcome. Suppose the agent's internal states to have objective tendencies of some determinate measure to cause certain volitional or actional outcomes. While this provides an opening in which the agent might freely select one option from a plurality of real alternatives, it fails to introduce a causal capacity that fills it. And what better here than it's being the agent himself that causes the particular action that is to be performed? It is this perceived inadequacy of attempts to identify an agent's free determination or control of his choices in event-causal or noncausal terms that motivates agent causationists to take it as an ontological primitive.

II Ontological and Metaphysical Commitments of any Agent Causal Theory

Many philosophers approvingly cite P.F. Strawson's (1962) charge that libertarianism involves "obscure and panicky metaphysics," and they often have the agent causal account of human freedom particularly in view. We will discuss shortly the nature of the posited agent-causal relation itself. For there to be such a relation, it seems that three further more general metaphysical theses must be true.

Agents as Substances Which Endure Through Time

Philosophers think of the persistence of objects through time in two opposing ways. On the 'temporal parts' ontology, an object persists by being composed of instantaneous temporal parts, in much the way that they have spatial parts. According to this view, just as my left foot is but a spatial part of me (and when we say that I exist there, we actually

mean that a part of me is there, that I overlap that region), so also the present stage of my existence is a temporal slice of my whole being, a component of the four-dimensional object spanning some eighty (?) years that, speaking tenselessly, is me. Clearly, a temporally extended object would not be suited to play the role of an agent cause of ever so many particular episodes in its own life. But neither are any of the momentary stages suitable, as these are not distinct from total states of the object at a particular time, and agent causation is supposed to be different from causation by states or events within the agent. Hence, there is nowhere to 'put' agent causation within the temporal parts theorist's ontology. Thus, agent-causal theories require that we think of objects (and agents specifically) as things that endure through time, such that they are wholly present at each moment of their existence, without being identical to the total state they are in at that moment.

Agents as Compositionally Irreducible Substances (Though Possibly Physically Composed)

It is highly plausible that, at any particular time, the being and activity of many sorts of composite objects is wholly constituted by the being and activity of their fundamental constituents. In such cases, the token causal relations into which the composites enter are not other than, but are instead wholly constituted by, the causal relations involving their current constituents. If this is true of human activity as well, then we do not engage in an ontologically distinctive kind of causation, as the agent causal theory supposes. So the agent causationist requires an ontology on which human persons are not only temporally

enduring but also synchronically irreducible substances—entities that are in some robust sense more than the sum of the constituents of their bodies.

Some agent causationists have supposed that human persons are simple (partless) substances (Reid, 1788; Chisholm, 1976). But others are quite emphatic that human persons are simply living animals, having no immaterial parts (Richard Taylor, 1966: 134-38; Randolph Clarke 1993: 201, n.14). How might human beings be composite but compositionally irreducible? O'Connor (2000: Ch.6) proposes that agent causal power and its allied properties are ontologically emergent, while still being powers and properties of the biological organism. On that account, a state of an object is emergent if it instantiates one or more simple, or nonstructural, properties and is a causal consequence of the object's exhibiting some general type of complex configuration. A property is 'nonstructural' just in case its instantiation does not even partly consist in the instantiation of a plurality of more basic properties. For detailed exploration of the individuation of emergent systems, see O'Connor and Jacobs (2003).

Causal Antireductionism

Finally, agent causationists must reject theories that purport to reductively identify the relation of causation with noncausal facts, such as certain patterns of actual similarity among event types, as on the traditional Humean analysis, or of counterfactual similarity, as on David Lewis's neo-Humean view (1986b). Agent causation, understood as a kind of single-case control functioning more or less independently of the agent's dispositional states, clearly cannot be understood in any such terms. Thomas Reid (1788) and George Berkeley (1710 and 1713), the two most prominent defenders of agent causal theories in

early modern philosophy, went so far as to hold that agent causation is the only form of causation properly so-called. The regular patterns exhibited in our experience among sensible objects are directly produced by God, the supreme agent cause. Contemporary theorists, by contrast, develop their view of agent causation as an extension of more general anti-reductionist conceptions of causation, as we shall now observe.

III Theories of Agent Causation

Agent causationists suppose that there is an ontologically primitive causal relation between agents and (some part of) their free actions. In order to assess the viability of this claim, we need to answer a number of questions concerning the details of the proposal. How are agent causation and event causation related? What precisely is it that the agent causes? How do the agent's reasons explain an agent-causal action? And under what circumstances might agent causation occur? Agent causationists have answered these questions in different ways.

Agent Causation and Event Causation: Their Distinctness and Their Relationship

Contemporary agent causationists maintain that agent and event causation are equally basic, related features in the natural order of things. Clarke's point of departure is the novel analysis of event causation proposed independently by Fred Dretske (1977), David Armstrong (1983), and Michael Tooley (1977, 1987). In basic outline, the view identifies laws of nature with certain primitive, contingent, and second-order relations among universals, ones that are specified as satisfying certain theoretical requirements associated with our concept of scientific law, structuring the distribution of event tokens in

accordance with certain of their types. The event causal relation, conceived as a type, is a special subset of these and is instanced between first-order events.

Clarke proposes that the very relation of causation that is thus theoretically identified within the domain of complex universals also holds between agents and their actions in instances of freely performed action. The sole differences between event and agent causation are the causal relata and the form of the laws structuring the distribution of their instances. (Furthermore, it is at most a contingent truth that the only causation by substances involves purposive agents. According to Clarke, if agent causation is possible, so is, e.g., rock causation.) Event causings are, at least as a contingent matter of fact, structured by probabilistic or deterministic laws. Clarke argues that, given what we know about human agency, the agent causationist should likewise to maintain that agent-causal relations are law governed. For example, we might suppose there are laws of nature to the effect that the causal relation obtains between agents and certain events only where agents have properties required for reflective practical reasoning and the caused events are instances of acting for reasons. Further, it might be a law of nature that whenever agents with such capacities do act on reasons, the causal relation obtains between the agent and the action (though the laws and antecedent circumstances do not imply which action will be so caused.)

O'Connor (2000, 2008) develops an account of agent causation in terms of the older, power-based conception of causation that is currently receiving renewed attention (on which, see Molnar 2003, Bird 2007, Jacobs 2007, and the essays in Handfield 2008). On this conception, natural properties are irreducibly dispositional: they are (or are closely associated with) tendencies to interact with other qualities in producing some effect, or

some range of possible effects. These dispositions may be probabilistic, such that there are objective probabilities less than one that a cause will produce its characteristic effect on a given occasion, with deterministic propensities simply being a limiting case where the probability is one. Some will endorse the possibility of pure, unstructured tendencies, ones that are nondeterministic and yet have no particular probability of being manifested on a given occasion.

O'Connor contends that agent causation (whether actual or merely possible) is an ontologically primitive type of causation, one that is uniquely manifested by (some possible) persons and is inherently goal-directed and nondeterministic. It is not directed to any particular effects, but instead confers upon an agent a power to cause a certain type of event within the agent: the coming to be of a state of intention to carry out some act, thereby resolving a state of uncertainty about which action to undertake.

What Does the Agent Cause?

We have just noted that O'Connor thinks of agent causation as directed to immediately executive states of intention. The agent's causing such an intentional state is the agent's forming of a choice or decision, and it is also the agent's basic action, typically the initiation of an extended sequence of event-causal processes constituting a wider observable action. (For broadly similar views, see Campbell 1967, Zimmerman 1984, and Donagan 1987.) Chisholm (1966) also thought of agent causation as a kind of action-triggering event, though, in keeping with his repudiation of choice or volition as a basic mental category, he supposed that what is caused is a nonintentional, neurophysiological event. Finally, while Clarke does speak of agents as making choices, he says that agents

cause their entire actions (as does Taylor). But since this suggestion is perhaps difficult to make out and is not essential to Clarke's development of an agent-causal account, we will assume in what follows that agents cause action-triggering, intentional events.

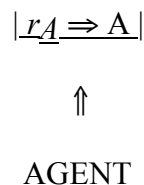
Agent Causation and Reasons-based Explanation of Action

In the distant past and continuing up to fairly recent times, agent causationists (e.g., Reid and Taylor) have tended to construe reasons explanations of agent-caused actions as irreducibly and purely teleological. Reasons provide explanations for actions not by causally contributing to their occurrence but by identifying the goal for the sake of which the action is undertaken. O'Connor (1995) proposes that the noncausal explanatory link between prior motivational states and agent-caused actions is forged via a distinctive type of content in the action-initiating intention that the agent causes. The content of these intentions is not merely that one perform an action of type A, but that one perform an action of type A *in order to satisfy reason r* (e.g., a prior desire or intention). If intentions have this rich sort of content, then—to answer Davidson's challenge to noncausalists—the difference between acting to satisfy reason *r1*, acting to satisfy *r2*, and acting to satisfy both will be a function of the content of the intention that one cause to occur, a content that will differ in each of these cases. In deciding which action one will undertake, one is inter alia deciding which reason one aims to satisfy. It is further required that this intention, once generated, causally sustains the completion of the action in an appropriate manner.

However, there is reason to find inadequate any noncausal account of the influence of reasons. Reasons move us more or less strongly toward certain courses of action varies

considerably, and no noncausal account of motivational strength is in the offing. (It is plainly not enough, for example, to maintain, with Chisholm 1985 and O'Connor 1995, that one's having a reason to A is a necessary causal condition on one's now causing the intention to A.)

Randolph Clarke develops his account of agent causation with the explicit goal of remedying this defect in the traditional picture. According to the earliest version (Clarke 1993), when an agent acts freely, her coming to have reasons to so act (r_A) indeterministically cause her action A. The agent figures into the picture by causing, not the action simpliciter, but the action's being done for those reasons. If we let " \Rightarrow " stand for the causal relation, we may diagram the basic picture thus:



One worry with this picture is that while it allows for reasons having varying strength—now explicated as a measurable causal tendency to produce an action—it's not clear that their having the strength they do influences the agent's activity. To be sure, Clarke says that the reasons indeterministically cause the *action* itself. But this, for him, does not include the agent's causal activity. And this seems odd: is not my directly causing some outcome something I do? And in any case, don't we want to say that my reasons have varying degrees of influence over this causality, whether or not we conceive it as part of my action? The above account doesn't make clear how this might be.

Addressing these concerns, Clarke (1996) maintains instead that the agent and the agent's indeterministic state of having reason r_A *jointly* produce A. The agent's causal capacity consists in the ability to make effective an indeterministic propensity of one's reason to bring about A, not by directly producing a causal relation between two events, but in the sense of 'acting alongside' or bolstering the tendency (whether it be of a low or high probability measure), ensuring that it will achieve its characteristic effect.

$$\begin{array}{l} r_A \quad \} \\ + \quad \} \Rightarrow A \\ \text{AGENT } \} \end{array}$$

Does this achieve the desired integration of my tendency-conferring reasons and my agent causality? It may seem that it does not. In any given instance, the action has some chance of occurring (and on occasion will occur) apart from the agent's activity—else what is meant by saying that the reason has a tendency to produce the action? This would be to conceive reasons as actively *competing* with the agent, qua agent. But Clarke says something further here: "...suppose that, in the circumstances, whichever of the available actions the agent performs, that action will be performed, and it will be caused by the reasons that favor it *only if* the agent causes that action" (1996: 25, emphasis added).

The proposed linkage is developed further in Clarke (2003). In the presence of a 'live' agent causal capacity and a plurality of motivational reasons $r_1 \dots r_n$ having nonzero probabilities of causing actions $A_1 \dots A_n$, it is a law of nature that:

- i) whatever action is performed will be caused by the agent,

ii) a reason r will cause an action A only if the agent causes it, and

iii) the agent will cause an action only if a corresponding reason causes it.

(We do not collapse clauses (ii) and (iii) into a biconditional since we need not say that in every case there is some reason such that *it* will cause action A if the agent does.)

O'Connor (2000), working within the causal powers approach to causation, augments his (1995) account by suggesting that an agent's coming to recognize a reason to act induces or elevates an objective propensity for her to initiate the behavior, and motivational strength consists in the strength of this propensity. This allows us to say that agent causation is a probabilistically *structured* capacity. It will be structured not only by tendency-conferring states of having reasons to act in specific ways but also by more enduring states of character, involving relatively fixed dispositions and long-standing general intentions and purposes around which her life has come to be organized. As on earlier accounts, the agent is the sole causal factor directly *producing* her intention to A (not a co-cause along with her reasons, as on Clarke's view), but her deliberation and activity take place within an internal context (including her total motivational state) that has probabilistically-delineated causal structure.

O'Connor (2005, 2008) further distinguishes acting *on* a reason and the stronger notion of acting *for* a reason. An agent S freely acts on a reason r in causing the occurrence of intention i just in case S causes the occurrence of i and r is a structuring cause of his so doing, altering the prior probability of his so acting. Note that here the causal influence of the reason might be entirely unconscious, and where it is, one might suppose that the agent's freedom is diminished to some extent. However, it seems that there are other cases where I not only am conscious of certain reasons that favor the

course of action I am choosing but expressly choose the action for the purpose of achieving the goal to which those reasons point. This goal enters into the content of the intention I bring into being. In such cases, I act on a prior desire or intention *r* and I cause an intention with a certain type of content, *A-ing for the sake of G*, where G is the goal of the consciously-grasped reason *r* for which I act. Now, since I freely and consciously bring the intention into being and thus give it just this purposive content, that purpose cannot but be one for which I am acting. What is more, a further explanatory connection between that reason and the choice is forged beyond the reason's influence on the choice's prior probability. This connection consists in the conjunction of the external relation of prior causal influence and the purely internal relation of sameness of content (the goal G). There may be several reasons that increase the likelihood that I would cause the intention to A. In the event that I do so, each of these reasons are ones *on* which I act. But if I am conscious of a particular reason, *r*, that promotes a goal G (and no other reason promotes that goal), and I cause the intention to A for the sake of G, then *r* plays a distinctive explanatory role, as shown by the fact that it alone can explain the goal-directed aspect of the intention's content. It alone is one *for* which I act.

IV Objections to Agent Causal Theories

I now consider several objections to agent causation. The first four offer reasons for thinking that there is a fundamental conceptual flaw or weakness in the picture of free action offered by the agent causationist, a sufficiently serious flaw or weakness as to call into question its coherence. The remaining three argue that, even assuming that agent causation is coherent, appeal to agent causation, or to agent causation under certain

assumptions made by its proponents, is useless in trying to reconcile free will with indeterminism.²

The 'Actions Are Datable, Agents Are Not' Objection

C.D. Broad famously objected to the possibility of agent causation as follows:

I see no *prima facie* objection to there being events that are not completely determined. But, in so far as an event *is* determined, an essential factor in its total cause must be other *events*. How can an event possibly be determined to happen at a certain date if its total cause contained no factor to which the notion of date has any application? And how can the notion of date have any application to anything that is not an event? (1952: 215)

Broad's objection, or something like it, would have considerable force against an agent-causal view that maintained that nothing about the agent at the time of his action was explanatorily relevant to its performance. Such an "action" would indeed seem freakish, or inexplicable in any significant way. But no agent causationist imagines such a scenario. On the contemporary accounts we've noted above, the agent's capacity to cause action-triggering events is causally structured by the agent's internal state, involving the having of reasons and other factors, before and up to the time of the action. These events within the agent suffice to explanatorily ground the agent's causing the event to happen "at a certain date" without collapsing the view into one on which those events themselves produce the action.

Randy Clarke (2003: 201-2), an erstwhile defender of an agent-causal account of freedom, has recently claimed that a modified version of Broad's objection has some force. Events, but not substances, are 'directly' in time in that their times are constituents of the events. By contrast, he maintains, "a substance is in time only in that events involving it...are directly in time." (This is supposed to be directly parallel to a reverse contention with respect to space, on which substances occupy space directly whereas events in their careers occupy a location only via its constituent object.) From this, he suggests, one can argue that the fact that effects are caused to occur at times "can be so only if their causes likewise occur at times—only, that is, if their causes are directly in time in the way in which events are but substances are not." (2003: 201)

The contention that drives this argument is obscure. It can easily be taken to suggest that events are ontologically more fundamental than objects, a contentious claim that any agent causationist will reject out of hand. But if this is not being claimed—as the reverse contention regarding occupation of space confirms—the point is unclear. What does it mean, exactly, to say that an object exists at a time "only in that" events it undergoes exist at that time? It cannot be the claim that the object's existing at that time metaphysically depends on the event's existing, as the object might have undergone another event at that time instead. If we weaken the claim to the plausible observation that, necessarily, an object *O* exists at time *t* only if there is some event or other involving it that occurs at *t*, the dependence is no longer asymmetrical: for any event occurring at *t* that involves an object, *O*, necessarily, *that* event exists at *t* only if *O* exists at *t*. Since I can think of no other way of explicating the 'exists only in that' relation, I do not see here

a promising basis for Broad's assertion that the cause of an event can only be another 'datable' entity.

The 'Uniformity of Causal Power' Objection

A second objection on which Clarke (2003) puts a great deal of weight begins with the following observation. If there is such a thing as agent causation, then there is a property or set of properties whose dispositional profile is precisely to confer on the agent a capacity to cause an intention to act. Notice how this contrasts with other causal powers in a very basic respect: the obtaining of properties that constitute 'event-causal' powers themselves tend towards certain effects (conditional on other circumstances). Hence,

Event-causal powers are tendencies towards effects, i.e., the powers themselves are disposed to produce effects.

Agent-causal power confers a capacity upon agents to produce effects, i.e., the power is not disposed to produce anything, it merely confers on its possessor a generic disposition to cause effects.

The uniformity objection to the thesis of agent causation is simply that it is doubtful that there can be any such property that fundamentally "works differently" (by conferring a power on its possessor to cause an effect). (2003: 192-3) If true, "causation would then be a radically disunified phenomenon" (2003: 208), and this is evidently a bad thing.

We may read this objection as making the claim that the ontological category of *property* has an abstract functional essence that includes the tendency in the presence of other properties towards the direct, joint production of certain effects. Is there reason to think that this is so? Better, assuming that it is so, at *what level* of abstraction should the

thesis be applied? Consider that, in the advent of statistical laws in fundamental physics, many metaphysicians are now comfortable with the notion that there are nondeterministic dispositions varying in strength along a continuum, with deterministic potentialities merely being a limiting case. Consider further that, while properties typically work in tandem towards effects, a natural way of interpreting the phenomenon of radioactive particle decay is as an entirely self-contained process whose timing is radically undetermined by any sort of stimulus event. Finally, some adhere to the truth of (and still others to the possibility of) a view that all or many conscious mental properties are intrinsically intentional while this is true of no physical properties. None of these claims concern free will, and yet all posit a kind of variability in the nature of dispositional properties that warrants classifying them into different basic types. Given these examples, it is hard to see why there may not be a further partition of types of the sort envisioned by the agent causationist. Doubtless there is a unity across these divisions at *some* level of abstraction. But assuming the agent causationist's position is otherwise motivated, he may reasonably contend that it must be sufficiently abstract as to encompass the division his theory requires. Indeed, why may not the unity of *basic* dispositional properties simply consist in their making a net addition to the pool of causal powers?

The 'Self-Creation is Impossible' Objection

Galen Strawson (1986: Ch.2; 1994) gave vivid expression to another old objection to indeterminist views of freedom, one that goes back at least as far as Leibniz. He argues that such views (unwittingly) entail an infinite regress of choices corresponding to every indeterministic choice. Since how one acts is a result of, or explained by, 'how one is,

mentally speaking' (*M*), for one to be responsible for that choice one must be responsible for *M*. To be responsible for *M*, one must have *chosen to be M* itself — and that not blindly, but deliberately, in accordance with some reasons *r1*. But for that choice to be a responsible one, one must have chosen to be such as to be moved by *r1*, requiring some further reasons *r2* for such a choice. And so on, ad infinitum. Free choice requires an impossible regress of choices to be the way one is in making choices.

The non-agent causationist Alfred Mele (1995: 221ff.) argues that Strawson misconstrues the locus of freedom and responsibility, by the lights of just about any theorist (including compatibilists). Freedom is principally a feature of our actions, and only derivatively of our characters from which such actions spring. The task of the theorist is to show how one is in rational, reflective control of the choices one makes, consistent with their being no freedom-negating conditions. This seems right, although the agent causationist is likely to add that when considering those theories (both compatibilist and incompatibilist) that make one's free control to directly reside in the causal efficacy of one's reasons, it's entirely appropriate to worry about how I got that way in the first place. (Which is just to say, Strawson's argument when directed against such accounts is best understood as challenging the adequacy of their understanding of free control over one's choices.)

But let us consider what the agent causationist might say in reply to Strawson. Aware of certain reasons pro (*r1*) and con, I cause an action-initiating intention to A. This is explained by my having been aware of reason *r1* while deliberating and as I completed the action, a reason that increased the prior probability of my choosing to A. I did not directly choose to be in a state of being aware of and motivated by *r1*. I simply found

myself in that state, among others, and proceeded to deliberate. The totality of such conative and cognitive states circumscribed the range of possibilities for me, and also presumably the scope of responsibility *directly* connected to my free choice. But that choice was neither fully causally determined by those states nor merely a 'chancy' outcome of tendencies of those states. Instead, *I* directly determined which choice within the available range would be made. This choice is *explained by* 'how I was, mentally speaking,' at that time, but it is not fully a *result of* that state. These two factors are treated separately, on agent causal accounts (as Clarke 1997 observes, in discussing Strawson), permitting direct control of an action that is not 'blind.' I chose for certain reasons, but I was not constrained to do so; given that this is so, there is no need for me to have first freely chosen which reasons I would act upon.

Of course, there is a residual worry hinted at by Strawson's argument. We enter the world with powerful and deep behavioral and attitudinal dispositions. Long before we mature to the point of making sophisticated, reflective choices, we are placed in environments that mold and add to those dispositions. These factors heavily influence the early choice we make, even if they do not causally determine all of them. They certainly do determine that Billy will choose from only a very limited range of options in any given situation, a range that will differ quite a bit from that open to Susie in similar circumstances. These choices and continuing contingencies of circumstance, in turn, will sharply circumscribe the options Billy considers at a more reflective stage, when we begin to hold Billy accountable for his actions. The worry, then, is that factors unchosen by Billy largely account for the kinds of deliberation and the overall pattern of outcomes

of Billy's mature choices. Even if an agent-causal capacity is at work in these choices, Strawson might ask, is it autonomy enough?

Surely one must concede in response that responsibility for 'shaping who I am' and for the choices that ensue from this comes in degrees and, indeed, can only sensibly be measured within a limited scope of possibilities. We cannot hold Billy responsible for failing to consider an option entirely outside the range of his experience. And his responsibility for passing by options which are within the range of his experience but which he has had precious little opportunity to consider as attractive is attenuated. In concrete cases, given limited information, we hazard rough guesses on these matters. When we are confronted with an individual who quite deliberately and unhesitatingly makes a grossly immoral choice—indeed, who seems not to even consider the obvious moral alternative—the question one needs to ask is this: was there a point earlier in her life when paths were open to her (ones for which at each step of the way she had some significant motivation to pursue, and which she recognized as having moral significance) such that had she taken them she would now be such as to see the force of the moral considerations at hand? How 'difficult' would it have been for her to pursue such a path? Our guesses about such matters are exceedingly rough, and rely on the assumption that most mature individuals have a certain measure of rough moral sensitivity. Absent compelling information to the contrary, then, we deem it appropriate to hold individuals responsible for their own moral indifference.

These reflections open up large topics that cannot be considered in this essay: In what ways do moral responsibility and perhaps freedom itself come in degrees? Is there a coherent ideal of perfect responsibility and freedom, and if so, is the notion continuous

with the imperfect varieties manifested by ordinary human beings? For discussion, see O'Connor (2005, 2009).

The 'No-Explanation' Objection

The objection to which I now turn is old and familiar. Though it is directed at indeterminist accounts of the will generally, it lies in the background of some of the more recent objections, which we will subsequently consider, to the philosophical *usefulness* of the concept of agent causation.

The no-explanation objection begins by noting that if an event A1 is causally undetermined, there is at least one event type A2 that might have been instanced instead of A1—its obtaining had a nonzero causal probability in the total set of circumstances. This implies that any actual set of circumstances *C* obtaining prior to A1 that one might cite in a putative explanation of A1 was consistent with the occurrence of the alternative, A2. But then it seems to follow that *C* cannot explain why A1 obtained *rather than* A2. Finally, the conclusion is drawn that *C* cannot, after all, fully explain the occurrence of A1 itself, since to fully explain why an event occurred is *inter alia* to explain why it occurred rather than any alternative.

This objection, while quite common, is misguided. Not all causal explanations of events must be contrastive or imply the availability of contrastive explanations, for every possible contrast. (To explain X, one need not explain why X rather than Y, for every possible Y.) 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. There *is* an explanation to be had, just not a contrastive one. A certain type of contrastive explanation is available for deterministic systems only.³

The 'Rollback' and 'Luck' Objections

Peter van Inwagen (2000) contends that “even if agent causation is a coherent concept and a real phenomenon, and we know this, this piece of knowledge will be of no use to the philosopher who is trying to decide what to say about free will” (11). Specifically, it is of no use in showing “that—despite appearances—free will is incompatible with indeterminism” (11). His principal argument for this contention has come to be known as the “Rollback” argument.⁴ He asks us to imagine an agent, Alice, faced with a choice between lying and telling the truth, who tells the truth while having been able to lie. Consistent with the agent causal picture of freedom, we suppose that the choice was causally undetermined, such that each of the two outcomes had a significant probability of occurring, and that the choice was an agent-causal event—the agent’s causing the formation of an intention to tell the truth, which in turn caused the completion of the act. Now, van Inwagen says, suppose that God were to ‘roll back’ the universe to a moment shortly before the choice and allowed it to play itself out again. The outcome may have been the same as the original, but it also may have been different—there is no fact of the matter about what it *would* have been, just what the possibilities and their associated

probabilities are. And now suppose further that God enables 1,000 replays. Since the probabilities in each case would be the same, if we observed the whole sequence, we would very likely observe a convergence around a certain ratio in the distribution of outcomes—50/50, 30/70, or something else, as the case may be. We would ineluctably be led to conclude, van Inwagen suggests, that what happens in a given case—e.g., “its coming to pass that Alice agent-causes the truth-antecedent cerebral events”)—is “a mere matter of chance,” such that this would not appear to be a free act. But the lesson we learn applies even when, as in the actual case, there is no rollback, and each exact situation type occurs but once. The rollback scenario was merely a heuristic, designed to get us to see that appeal to agent causation does nothing to dispel the sense that causally undetermined outcomes are inherently a “mere matter of chance.” Any solution to the problem of freedom and indeterminism, therefore, must be found elsewhere.

The ‘luck’ objection invites us to contemplate, not intra-world identical undetermined choice situations obtained via rollback (a metaphysically dubious notion, it should be said), but inter-world cases. We imagine Alice and a counterpart Alicia in an identical world up to the moment of choice, such that Alice tells the truth and Alicia lies, and again we tell the story in a manner consistent with the agent causal story. If the bravely truth-telling Alice is commended, and the deceiving Alicia goes on to be exposed and suffers a negative consequence, isn’t Alice just lucky? After all, there was *nothing* whatsoever about her right up to the moment of the choice that distinguished her from Alicia, and so nothing about her that made the difference. Each had the same propensity to lie and to tell the truth. The conclusion drawn is that neither agent controlled the way

their respective cases unfolded in such a way that it was up to her that she told the truth (lied). (For a statement of this argument, see Haji 2004.)

The agent causationist contends that both these objections fail to take seriously the concept of agent causation. It is conceived as a primitive form of control over just such 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. Alice's causing her intention to tell the truth 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.⁵ So wherein lies the luck? (For such a response, see Pereboom 2005, Clarke 2005, and O'Connor 2007).

The 'Strengthened Luck' Objection

Mele (2006, 2007) adds a challenging variation on the luck objection. He grants that the agent causationist can say that there is no luck with respect to the agent's *doing as he does*. There is a causal power in play that yields that outcome and thereby allows for a noncontrastive explanation of it.⁶ But notice that the causal indeterminist can say that much as well! (On the latter picture, it is the causal efficacy of the agent's reasons that indeterministically yield the outcome, and causal control for such a theorist just consists in such efficacy, provided the causal process is not subject to causal deviance or external manipulation.) And agent causationists generally do not dispute that causally indeterministic agents exercise a form of control over their choices. Even so, the agent causationist still sees an ineliminable luck problem for causal indeterminism. Luck attaches not to the simple outcome itself, but to the *contrastive* fact that the outcome

occurred rather than the alternative. (In our example, it is lucky for Alice that she told the truth *rather than* lied.) Mele writes:

...if the question why an agent exercised his agent-causal power at *t* in deciding to *A* rather than exercising it at *t* in any of the alternative ways he does in other possible worlds with the same past and laws of nature is, in principle, unanswerable...because there is no fact or truth to be reported in a correct answer...and his exercising it at *t* in so deciding has an effect on how his life goes, I count that as luck for the agent.

(2006: 70)

Suppose we take this as a stipulative account (or sufficient condition) on luck *as Mele understands the notion*. If so, it is open to the agent causationist to deny that luck in this stipulated sense is of any significance whatsoever—not, for example, being relevant to freedom and moral responsibility. Mele in fact agrees! He does not press his luck objection as a deep skeptical worry about indeterministic freedom (as van Inwagen does with his Rollback argument). Instead, he wields it to neutralize the agent causationist's objection to causal indeterminism. Mele is a causal theorist of action, an agnostic about the truth of compatibilism, and one who believes we have freedom and certainly that it is possible. His aim is to show that the agent causationist gains nothing by positing a distinctively agential capacity of control.

In reply, Clarke (2005) argues that an agent causal capacity would provide a *stronger variety* of control than is available on causal indeterminism. On the causal indeterminist account, the outcome in a given case will proceed from one or another of the agent's own motivationally-grounded dispositions. As such, the outcome is plainly

controlled by the agent in a clear sense, much as any sophisticated regulational device that operated nondeterministically would control its own outputs. But what is lacking in each case is a certain enhanced *kind* of control: one that would be exercised were the agent in a maximally direct manner to bring about one or another option—to settle which of her probabilistic dispositions will be manifested on that occasion. The absence of this form of control, the agent causationist argues, is what grounds the original ‘luck’ charge against causal indeterminism. On that charge, luck attaches not to a contrastive fact about what occurs, but to the occurrence itself.

The ‘Probabilistically-Governed Agent Causation is Insufficient for Freedom’ Objection

The agent causationist takes it to be a virtue of her theory that it enables her to avoid a ‘problem of luck’ facing other indeterministic accounts. Agent causation is precisely the power to directly determine which of several causal possibilities is realized on a given occasion. However, Derk Pereboom has recently argued that this is so only if agent causation does not conform to pre-given indeterministic tendencies. He writes:

...to answer the luck objection, the causal power exercised by the agent must be of a different sort from that of the events that shape the agent-causal power, and on the occasion of a free decision, the exercise of these causal powers must be token-distinct from the exercise of the causal powers of the events. Given this requirement, we would expect the decisions of the agent-cause to diverge, in the long run, from the frequency of choices that would be extremely likely on the basis of the events alone. If we nevertheless found conformity, we would have very good reason to believe that the agent-causal power was not of a different

sort from the causal powers of the events after all, and that on the occasion of particular decisions, the exercise of these causal powers was not token-distinct.

Or else, this conformity would be a wild coincidence...(2005: 246)

Though Pereboom expresses the matter in epistemological terms, I take it that he intends to be making a linked pair of metaphysical claims, as follows. If agent-causal power is to enable the agent directly to determine which causally-possible choice obtains, and so overcome the luck objection plaguing other accounts of freedom, then it must be a different sort of power from the event-causal powers grounded in the propensities of one's reasons, such that its exercise is token-distinct from the exercise of any of these event-causal powers. And the latter condition can be met only if the outcomes of agent-causal events are not strictly governed by the propensities of any relevant set of obtaining event-causal powers.

The agent causationist readily endorses the first of these conditionals, on a straightforward reading of "different sort of power" and "token-distinct exercise." After all, the view posits a fundamental, irreducible power of agents to form intentions. But the second conditional directly rejects the viability of any account on which agent causal power is probabilistically structured by reasons. Why does Pereboom assert it? His thought seems to be that if the event of one's having certain reasons along with other prior events ensure that one's choices will fit a certain pattern – more accurately, make the pattern-fitting likely, given a sufficiently large number of cases – , then one's supposed agent-causal power in choosing is at best a shadowy accompaniment to the event-causal power. In truth, it is no power at all, as it adds nothing to the mix of factors

already in play. With no authority to act on its own, its presence makes no discernible difference to what occurs in the aggregate. If it would be a matter of luck, beyond my direct control, which of my indeterministic propensities happens to be realized on any given occasion, were the causal indeterminist account correct, then adding the ability to ‘directly determine’ the outcome wouldn’t help if I am ineluctably constrained by those very propensities.

In reply, the agent causationist will insist upon the importance of the distinction between (the persisting state or event of one’s having) reasons *structuring* one’s agent-causal power in the sense of conferring objective tendencies towards particular actions and reasons *activating* that power by producing one’s causing a specific intention. Nothing other than the agent himself activates the agent causal power in this way. To say that I have an objective probability of 0.8 to cause the intention to join my students at the local pub ensures nothing about what I will in fact do. I can resist this rather strong inclination just as well as act upon it. The probability simply measures relative likelihood and serves to predict a distribution of outcomes were I to be similarly inclined in similar circumstances many times over (which of course I never am in actual practice). From the agent causationist perspective, the reason that the alternative, causal indeterminist view is subject to the luck objection is *not* that it posits objective probabilities to possible outcomes but that it fails to posit the kind of control needed directly to determine what happens in each case. After all, were the causal indeterminist picture modified so that agents’ choices were caused but not determined by appropriate internal states whose propensities, while nondeterministic, lacked definite measure, the problem of luck or control would remain. Again, that problem concerns not prior influence but the ability

directly to settle what occurs on the occasion of a causally undetermined outcome. The agent causationist's solution is to posit a basic capacity of just that sort, while allowing that the capacity is not situated within an indifferent agent, but one with evolving preferences and beliefs. Surely having preferences does not undermine control!

Bibliography

- Armstrong, David. 1983. What Is a Law of Nature?. Cambridge: Cambridge University Press.
- Berkeley, George. 1710. Principles of Human Knowledge.
 -- 1713. Three Dialogues Between Hylas and Philonous.
- Bird, Alexander. 2007. Nature's Metaphysics: Laws and Properties. Oxford: Clarendon Press.
- Bishop, John. 1989. Natural Agency. Cambridge: Cambridge University Press.
- Broad, C.D. 1952. "Determinism, Indeterminism, and Libertarianism," in Ethics and the History of Philosophy. London: RKP.
- Campbell, C.A. 1967. In Defence of Free Will & other essays. London: Allen & Unwin.
- Chisholm, Roderick M. 1966. "Freedom and Action." K. Lehrer, ed. Freedom and Determinism. New York: Random House.
 -- 1976. Person and Object. LaSalle: Open Court.
 -- 1985. "Self-Profile," in R. Bogdan (ed.) Roderick M. Chisholm. Dordrecht: D.Reidel.
- Clarke, Randolph. 1993. "Toward a Credible Agent-Causal Account of Free Will." Noûs 27: 191-203.

- 1996. "Agent Causation and Event Causation in the Production of Free Action."
Philosophical Topics 24 (Fall): 19-48.
- 1997. "On the Possibility of Rational Free Action." Philosophical Studies 88: 37-57.
- 2003. Libertarian Accounts of Free Will. New York: Oxford University Press.
- 2005. "Agent Causation and the Problem of Luck," Pacific Philosophical Quarterly
86 (3), 408-421.
- Davidson, Donald. 1963. "Actions, Reasons, and Causes," Journal of Philosophy, 60,
685-700. Reprinted in D. Davidson, Essays on Actions and Events. Oxford:
Clarendon Press.
- 1980. 'Agency', in Essays on Actions and Events. Oxford: Clarendon Press.
- Davis, Wayne. forthcoming. "The Causal Theory of Action," in Constantine Sandis and
Timothy O'Connor (eds.), A Companion to the Philosophy of Action. Oxford:
Blackwell Publishing.
- Donagan, Alan. 1987. Choice. London: Routledge and Kegan Paul.
- Dretske, Fred. 1977. "Laws of Nature." Philosophy of Science 44: 248-68.
- Ellis, Brian. 2001. Scientific Essentialism. Cambridge: Cambridge University Press.
- Feldman, Richard H. & Buckareff, Andrei A. 2003. "Reasons Explanations and Pure
Agency," Philosophical Studies 112 (2), 135-145.
- Goetz, Stewart. 2002. "Review of O'Connor, *Persons and Causes*," Faith and Philosophy
19, 116-20.
- Griffith, Meghan 2005. "Does Free Will Remain a Mystery? A Response to van
Inwagen," Philosophical Studies 124 (3), 261-69.

- Haji, Ishtiyaque. 2004. "Active Control, Agent-Causation, and Free Action,"
Philosophical Explorations 7(2), 131-48.
- Handfield, Toby (ed.). 2008. Dispositions and Causes. Oxford: Oxford University Press.
- Hiddleston, Eric. 2005. "Critical Notice of Timothy O'Connor, *Persons and Causes*,"
Noûs 39 (3), 541-56.
- Jacobs, Jonathan D. 2007. Causal Powers: A Neo-Aristotelian Metaphysics. PhD. Thesis,
 Indiana University.
- Kane, Robert. The Significance of Free Will. New York: Oxford University Press.
 -- 2005. A Contemporary Introduction to Free Will. New York: Oxford University
 Press.
- Levy, Neil and McKenna, Michael (2009). "Recent Work on Free Will and Moral
 Responsibility," Philosophy Compass 4(1), 96-133.
- Lewis, David. 1986. "Causation." Philosophical Papers, vol. 2. Oxford: Oxford
 University Press.
- Lipton, Peter. 1990. "Contrastive Explanation," in D. Knowles (ed.), *Explanation and its
 Limits*, Cambridge University Press, 1990, 247-266; reprinted in D. Ruben (ed.),
Explanation, Oxford Readings in Philosophy, Oxford University Press, 1993.
- Lowe, E. J. 2001. "Event Causation and Agent Causation," Grazer Philosophische
 Studien 61, 1-20.
 -- 2008. Personal Agency: The Metaphysics of Mind and Action. Oxford: Oxford
 University Press.
- Markosian, Ned. 1999. "A Compatibilist Version of the Theory of Agent Causation,"
Pacific Philosophical Quarterly 80 (3), 257-277.

- McCann, Hugh. 1998. The Works of Agency. Ithaca: Cornell University Press.
- Mele, Alfred. 1992. Springs of Action. New York: Oxford University Press.
- 1995. Autonomous Agents. New York: Oxford University Press.
 - 2006. Free Will and Luck. New York: Oxford University Press.
 - 2007. "Free Will and Luck: Reply to Critics" Philosophical Explorations, 10, 195-210.
- Molnar, George. 2003. Powers: A Study in Metaphysics. Oxford: Oxford University Press.
- O'Connor, Timothy. 1993. "Indeterminism and Free Agency: Three Recent Views." Philosophy and Phenomenological Research 53: 499-526.
- 1995. "Agent Causation." T. O'Connor, ed. Agents, Causes, and Events: Essays on Indeterminism and Free Will. New York: Oxford University Press.
 - 1996. "Why Agent Causation?" Philosophical Topics 24 (Fall): 143-158.
 - 2000. Persons and Causes: The Metaphysics of Free Will. New York: Oxford University Press.
 - 2005. "Freedom With a Human Face," Midwest Studies in Philosophy, 29, 207-227.
 - 2007. "Is It All Just a Matter of Luck?" Philosophical Explorations, 10, 157-161.
 - 2008. "Agent-Causal Power," in Toby Handfield (ed.), Dispositions and Causes, Oxford: Clarendon Press, 189-214.
 - 2009. "Degrees of Freedom," Philosophical Explorations 12 (2), 119-125.
- O'Connor, Timothy and Churchill, John Ross. 2006. "Reasons Explanation and Agent Control: In Search of an Integrated Account," Philosophical Topics, 32, 241-254.
- O'Connor, Timothy and Jacobs, Jonathan. 2003. Philosophical Quarterly, 53, 540-555.

- Pereboom, Derk. 2001. Living Without Free Will. Cambridge: Cambridge University Press.
- 2004. "Is Our Concept of Agent-Causation Coherent?" Philosophical Topics 32, 275-86.
- 2005. "Defending Hard Incompatibilism," Midwest Studies in Philosophy 29, 228-47.
- 2007. "On Alfred Mele's *Free Will and Luck*," Philosophical Explorations 10 (2), 163-172.
- Pink, Thomas. 2004. Free Will: A Very Short Introduction. Oxford: Oxford University Press.
- Reid, Thomas. 1788. Essays on the Active Powers of Man.
- Rowe, William. 1991. Thomas Reid on Freedom and Morality. Ithaca: Cornell University Press.
- Schlosser, Markus E. (2008). "Agent-Causation and Agential Control," Philosophical Explorations 11 (1), 3-21.
- Sehon, Scott. 1995 Teleological Realism. Cambridge, MA: MIT Press.
- Stout, Rowland. forthcoming. "Deviant Causal Chains," in Constantine Sandis and Timothy O'Connor (eds.), A Companion to the Philosophy of Action. Oxford: Blackwell Publishing.
- Strawson, Galen. 1986. Freedom and Belief. Oxford: Clarendon Press.
- 1994. "The Impossibility of Moral Responsibility." Philosophical Studies 75: 5-24.
- Strawson, P.F. 1962. "Freedom and Resentment," Proceedings of the British Academy 48, 1-25.

- Swinburne, Richard. 1997. The Evolution of the Soul, rev.ed. Oxford: Clarendon Press.
- Taylor, Richard. 1966. Action and Purpose. Englewood Cliffs: Prentice Hall.
- Thorp, John. 1980. Freewill: A Defence Against Neurophysiological Determinism.
London: Routledge and Kegan Paul.
- Tooley, Michael. 1977. "The Nature of Law." Canadian Journal of Philosophy 7: 667-98.
-- 1987. Causation: A Realist Approach. Oxford: Clarendon Press.
- Van Inwagen, Peter. 2000. "Free Will Remains a Mystery," Philosophical Perspectives 14:1-20.
- Vesey, Godfrey and Flew, Antony. 1987. Agency and Necessity. Oxford: Blackwell.
- Widerker, David. 2005. "Agent-Causation and Control," Faith and Philosophy 22 (1), 87-98.
- Wilson, George. 1989. The Intentionality of Human Action. Stanford, CA: Stanford University Press.
- Zimmerman, Michael. 1984. An Essay on Human Action. New York: Peter Lang.

¹ An important recent contribution to indeterminist accounts of freedom is E.J. Lowe (2008). Despite Lowe's ostensible endorsement of a qualified version of agent causation (as part of a broader picture on which all causation is substance causation), his view has a closer affinity to contemporary versions of noncausalism and so I shall not treat it further here. Lowe defends a view on which free actions are initiated by uncaused volitions. He further contends that this is consistent with agent causation by saying that it is *by* the uncaused willing of the action that the agent causes his bodily movements. If Lowe were

to go on to say that volitions have internal causal structure of the form, *agent-S-causes-event-e*, then his view would indeed be a form of agent causalism. But he appears instead to think of volitions as simple mental occurrences, albeit ones that are ‘intrinsically active’ (McCann 1998) or are exercises of the (noncausal) ‘power of freedom’ (Pink 2004). And that is just what noncausalists such as the two authors parenthetically noted maintain, claiming that appeal to a primitive relation of agent causation is otiose.

² For a lucid response to some of these objections, the reader is directed to Clarke (2003: Ch.9). Clarke goes on, however, to advance objections of his own, a couple of which are discussed in the text. Some of the material in this section is taken from O’Connor (2008).

³ There is much more to be said about contrastive and noncontrastive explanations (including most importantly the fact that contrastive why questions—why X rather than Y—often seek a different kind of information than merely the sum of explanations for X and for not-Y) but exploring those matters fully would not affect the basic point made in the text. An excellent starting point in exploring these matter is Lipton (1990).

⁴ Van Inwagen seeks to bolster his argument by appeal to a second “Promise” argument. For reasons of space, and because I take it to have been decisively rebutted by Clarke (2003: 168-70), I won’t consider this further argument here.

⁵ Widerker (2005) and Kane (2005: 51) dispute the agent causationist’s contention that agent causation is transparently a variety of agential control. I believe this objection to be misconceived, or at least wrongly posed. The agent causationist takes agential control of a freedom-grounding sort as a primitive, both ontologically and conceptually. She then tries to motivate this posit by showing how one might integrate such a primitive feature of control within a wider system of concepts concerning causation, properties, guidance

by reasons and so forth. The positing of primitives is to be judged by their theoretical fruitfulness, not by whether it causes the ‘light of reason’ to glow as one contemplates it. For this reason, Widerker should either argue that the agent causationist does not succeed in integrating her primitive into a plausible understanding of these other concepts or argue that the effort is unnecessary, as we can make do perfectly well without it.

Note that the simple indeterminist makes a structurally similar move in contending that choices are simple events that are controlled by the agent in virtue of simple intrinsic qualities, variously described as “actish phenomenal quality” (Ginet, 1990), “spontaneity” (McCann 1998), or the (noncausal) “power of freedom” or “choice” (Pink 2004; Goetz 2002). The agent causationist’s objection to this move is not that the mere fact that it posits a primitive but that the posit conflicts with the deeply intuitive judgment that the idea of control is tightly bound up with the idea of causation. Positing that control is primitive *and noncausal* in nature is mysterious, and the agent causal theory seems to provide the necessary correction: control *does* flow from the intrinsic character of choice, but that character has causal structure.

⁶ Not every proponent of the luck objection against agent causation keeps the contrastive/non-contrastive issues distinct. See, for example, the way that Levy and McKenna (2009: 120-1) appear to slide from the unavailability of contrastive explanation, even on agent causation, to its being doubtful that an agent could have made an alternative choice “for reasons.”