Free agency and materialism

Article · January 1996

CITATIONS
19

READS
70

2 authors, including:

Jan Cover
Purdue University
34 PUBLICATIONS **606** CITATIONS

SEE PROFILE

3

Free Agency and Materialism

J. A. Cover and John O'Leary-Hawthorne

Famously, Kant insisted that, given that the natural order is deterministic, we must conceive of free agents as somehow standing outside that order. The determinism that Kant had in mind is familiar enough; it is one according to which the laws of nature and the past together render impossible all but one subsequent course of events. Our contemporary picture of the natural order is not as a deterministic order in this sense. So if this sort of determinism provides the only legitimate reason for placing agents outside the natural order, then naturalism about free agents is in very good shape.

Yet there is a different sort of determinism associated with naturalism that, while very much alive today, has not found its way into discussions of human freedom. It is a sort of determinism according to which the microphysical world determines the distribution of the higher level properties of material beings, adumbrated in various popular supervenience theses. Given this sort of determinism in the natural order, must we again conceive of genuinely free agents as somehow standing outside that order? We shall be addressing that question in this chapter.

Three Tenets of Mind/Body Materialism

Let us begin with three assumptions that are typically part of any materialist worldview that is not also eliminativist about thinkers and their mental lives.

Assumption 1: People are wholly material beings.

That is to say, human persons have no immaterial parts. Whenever any human person says or thinks "I think," that person refers to a material being.

> Assumption 2: Alien worlds aside, mental states supervene on microphysical states.[56]

That is to say, necessarily, if the distribution of microphysical properties at two possible worlds is exactly the same, then the distribution of psychological properties at those worlds will be exactly the same. The supervenience thesis of Assumption 2 should be distinguished from some stronger theses, such as

> 2A. The mental states of an individual supervene on the microphysical states of that individual, and
>
> 2B. The global distribution of mental states at a time supervene on the global distribution of microphysical states at that time.

Recent discussions of mental content have urged the externalist lessons that (i) owing to the fact that environment partially constitutes content, 2A is false, and (ii) owing to the fact that history partially constitutes content, 2B is false.[57] Thus, for example, there are possible worlds exactly like the actual world now but where a different individual is the referent of "Napoleon," and thus where a different *de re* proposition (hence a different belief) is expressed by "Napoleon was a man."

While supervenience theses 2A and 2B may be false, there may well be true supervenience theses, akin to them, that are restricted to a certain domain of psychological properties. For example, if there is such a category as "narrow content," then a supervenience thesis in the style of 2A restricted to narrow contents may well be acceptable.

Return again to Assumption 2. Why the qualification about "alien worlds"? Even most hard-nosed materialists will concede that there is a logically possible world that duplicates all the physical properties of this one but where there are extra immaterial entities—angels or spooks—thrown in. Assuming there are no angels or spooks at the actual world, the physical properties of that world will duplicate those of our own, but the mental properties will not. An unrestrained supervenience thesis according to which physical similarity between worlds guarantees psychological similarity is thus too strong: so-called "alien worlds," with fundamental ingredients that do not figure at the actual world, will have to be excluded insofar as one claims that physical similarity guarantees psychological similarity.[58]

Our third assumption offers another restricted supervenience thesis in the style of 2A:

> Assumption 3: Immaterial beings aside, the property of *an agent's making a free decision at t* supervenes upon the intrinsic microphysical history of an agent up to *t*.[59]

That is to say, if there are two possible material beings that are duplicates with respect to intrinsic microphysical history up to *t*, then both or neither makes a free decision at *t*. Now it must be granted that the *content* of an agent's free decision may not supervene on the intrinsic microphysical history of that agent. (The *de re* content of an agent's decision, say, "to study the life of Napoleon," will not supervene on the agent's current microphysical structure.) Nevertheless, Assumption 3 seems to us a rather plausible view to take for anyone accepting Assumption 2. For if the mental globally supervenes on the microphysical, then Assumption 3 could fail for only one of two reasons: (a) because the microphysical future is partially constitutive of whether one makes a free decision, or (b) because the current microphysical environment outside of the agent is partially constitutive of whether one makes a free decision. Option (a) seems outright implausible. Option (b) seems unlikely as well, once one remembers that freely making a decision to do *x* doesn't entail that one is able to do *x* (though, of course, it may well entail being able to decide not to do *x*). One can freely decide to leave the room even though one is not free to leave the room; one can freely come to the decision to pull the trigger even if the trigger is stuck or locked.

Now, Assumptions 1–3, even if accepted, still leave open a number of questions concerning the metaphysics of mind. Among them are

> (i) Is token dualism true? Are token mental events and states identical with or distinct from (though supervenient upon) complex microphysical states and events?
>
> (ii) Is property dualism true? Are mental properties identical with complex (presumably, hugely gerrymandered) microphysical properties or distinct from (though supervenient upon) complex microphysical states and events?

Our concern here is not to answer these questions, but instead to ask: How well do Assumptions 1–3 cohere with our ordinary conception of freedom and agency, coupled with our commonsense belief that we are

agents who are sometimes free? Note that Assumptions 1–3 do not entail determinism at the microphysical level. Thus, even if our ordinary conception of freedom is incompatible with the thesis that our actions are determined by the past and the laws of nature, it does not immediately follow that our ordinary conception of freedom sorts ill with Assumptions 1–3.

Assumptions 1 and 2 are, by our lights, obligatory for any materialist about the mind. It is unclear whether "token monism" and "property monism" are also obligatory. We don't want to pin too much on the materialist here; we recognize that some calling themselves materialists will balk at having property monism pinned on them. So we shall require only that the materialist be committed to Assumptions 1 and 2. We have suggested, however, that anyone subscribing to Assumptions 1 and 2 will also inevitably find Assumption 3 compelling. So we shall also presume that the materialist about the mind will be prepared to endorse 3. The main question of this chapter may thus be posed in the following way: What are the costs of wedding our ordinary conception of freedom with materialism about the mind? Is materialism about the mind compatible with our ordinary conception of freedom?

Before proceeding directly with these questions, it will be helpful to say a bit more about our ordinary conception of freedom.

Robust and Deflationary Accounts of Freedom

Let's distinguish robust from deflationary accounts of freedom. The robust theorist argues that there is a phenomenon out there in the world meeting the contours of our ordinary conception of freedom. The deflationary theorist argues that while there is something in the world deserving the titles "freedom" and "agency," nevertheless significant bits of the picture surrounding ordinary thought and talk about freedom and agency in fact have no correlate in the world, and thus are either to be dropped altogether from our best theory or else to be somehow accommodated but recognized for what they are. (There is also the eliminativist about freedom—who argues that ordinary ascriptions of freedom and agency are simply false or else truthvalueless.)

Standard compatibilist accounts of freedom seem deflationary at best. Not all compatibilists will agree. They may insist that their compatibilism does full justice to our ordinary conception of freedom—insisting, for example, that all it means to be free in the ordinary sense is for one's actions to be the causal upshot of one's beliefs and desires (it being

neither here nor there where one's beliefs and desires come from). A full treatment of this compatibilist claim to do justice to our ordinary conception of freedom is beyond the scope of this paper.[60] It will suffice for now simply to declare our own orientation—that any such claim on behalf of compatibilism is at best poor anthropology. There are real phenomena, such as the following, which need to be explained: When ordinary people come to consciously recognize and understand that some action is contingent upon circumstances in an agent's past that are beyond that agent's control, they quickly lose a propensity to impute moral responsibility to the agent for that action. We can readily explain this fact by supposing that ordinary people have a conception of freedom, agency, and moral responsibility according to which an action by an agent is free and accountable only if that action is not fully determined by circumstances, past or present, that are beyond the agent's control. Similarly, we believe that the best explanation for why so many philosophy students find compatibilism *prima facie* implausible is that they carry with them a workaday conception of freedom that cannot be done full justice by the compatibilist account.

It is somewhat easier to be sympathetic with standard compatibilist accounts if they are offered as explicitly deflationary theories. There is nothing incoherent in the nature of a claim to the effect that ascriptions of choice are often true even though certain philosophical pictures accompanying such ascriptions are confused. By way of analogy, it may well be argued that certain conceptions that we ordinarily have of time are strictly speaking wrongheaded and yet allow that there is nevertheless a phenomenon deserving the title "time." (One way of fleshing out a semantics along these lines is to say, with David Lewis, that a near-realization of folk theory about some subject matter may provide the denotation of the relevant folk predicates even though there is no perfect realizer of folk theory concerning that subject matter.[61]) Clearly, deflationary accounts of freedom are compatible with the materialist theory of mind. Our concern in this chapter is to inquire whether a robust account of freedom is so compatible, and if so, at what cost materialism is brought alongside robust freedom.

What will a robust conception of freedom look like? So-called agency theories attempt to provide the *bare bones* of a metaphysical story concerning robust freedom and moral responsibility. Here, in short, is what agency theory tells us:

(i) Among the things that exist, there are *agents*.

(ii) There is a fundamental relation between agents and actions—call it *agent*

causation—such that by standing in that relation to actions, agents count as performing those actions freely and count as being accountable for those actions.

Some have said that agency theory is outright incoherent, since it aims to find a logically impossible *via media* between statistically random actions (i.e., actions rendered less than inevitable by the past course of events and the laws of nature) and actions determined by the past and the laws of nature. We don't see the incoherence. There is nothing outright contradictory about the following claim: There is some differentium R such that the class of undetermined actions divides into those that are morally accountable and those that aren't, according to whether they have or lack R respectively. Perhaps it is supposed to be self-evident that if any undetermined event is morally accountable, every undetermined event is; or perhaps it is supposed to be self-evident that no undetermined event is one for which anyone could be morally responsible. But neither of these is self-evident. So agency theory is not outright incoherent.

Agency theory has at least two virtues relative to the aim of doing justice to our ordinary conception of freedom. First, agent causation, if true, offers the real possibility of *doing otherwise*. Consistent with the claim that agent *s* stands in *R* to action *a*, one can insist that *a* is not determined by past events or laws of nature; *s* could do otherwise than *a*. The second virtue addresses itself to the belief that it is *we* who act properly or blameworthily, and, moreover, that this is not simply a matter of there being certain properties we happen to be carrying around, but rather that some action springs from our being, so to speak—that the action in some unnegotiable sense comes from us.[62] In at least these two ways, then, it might be felt that agency theory does full justice to our ordinary conception of freedom, and that anything less would be tantamount to a deflationary view.

It is worth noting that someone might in effect offer a version of agency theory preserving one but not both of these virtues. Consider Leibniz, for example. According to Leibniz, one's actions flow from one's individual essence.[63] Thus, in a very deep and unequivocal sense, one's actions express who *one* is. Indeed, actions flowing from one's individual essence are certainly not contingent upon circumstances beyond one's control. For an action to be so contingent requires that there be possible worlds where the circumstances are different and where, as a result, one does something different. But if some trait or disposition or action were necessitated by one's individual essence, then quite clearly, that trait or disposition or action would not be so contingent. By

capturing the second virtue of agency theory, Leibniz thus preserves something quite intuitive. If I act badly in some way, then on the Leibnizian conception, I cannot very well "put that down" to bad upbringing or to some purely circumstantial or "extrinsic" facts, since the action itself reflects my very essence—who I am. It is not as if, had *my* upbringing being different, I would have been a better person. (It is important, further, to Leibniz, that the action stem from my individual essence and not my essence *qua* human being, since otherwise that action would be an expression of humanity in general rather than who I am in particular.) Clearly, this conception of freedom, while preserving the second virtue, goes rather a short distance toward preserving the first (nor do the various Leibnizian maneuvers around the notion of inclining without necessitating go very far here). We shall be focusing here on a fully robust conception of freedom, one that aims to preserve both strands, acknowledging that there may be middle ground (e.g., Leibniz) between that fully robust conception and familiar deflationary views.

The Mystery of Agent Causation

If we can suppose that agency theory goes some distance toward capturing a bare-bones story about our ordinary conception of freedom, we cannot pretend that the concept of agent causation approaches anything like a clear and distinct idea. Agency theory is radically underdeveloped and seems likely to remain so. The point can be best illustrated, in the present context, by examining the most recent attempt to develop a theory of agent causation, found in Randolph Clarke's "Toward a Credible Agent-Causal Account of Free Will."[64] Clarke makes two basic moves in attempting to render agent causation more palatable. The first—which we applaud—is to suggest that while facts about an agent's prior reasons for action will not be determining causes in cases of agent causation, they can nevertheless properly be counted as causally relevant to the freely produced action. The second basic move is to offer the following story about the agent-causal relation:

> What remains is to say just what this relation is. The prevailing tendency among agent causalists and their critics alike on this point has been to stress how different agent causation is from event causation and indeed how 'mysterious' the former is. However, the proper line here, I believe, is to maintain that agent causation, if there is such a thing, is (or involves) exactly the same relation as event causation. The only difference between the

two kinds of causation concerns the types of entities related, not the relation. (197)

Clarke then goes on to recommend that we take

> the causal relation to be among the basic constituents of the universe. Causation may be held to be a real relation between particulars, one that, although analyzable, is not reducible to noncausal and non-nomological properties and relations. . . .
>
> One type of realist account of event causation can be sketched, in broad strokes as follows. An event (particular) causes another just in case the relation of causation obtains between them. Two events can be so related only if they possess (or are constituted by) properties that are in turn related under a law of nature. Ultimately, then, causal relations are grounded in laws of nature, which consist of second-order relations among universals.
>
> Such an account resembles that favored by Tooley for event (or, as he would have it, state-of-affairs) causation. Tooley maintains that the relations involved in this sort of account—causation, as well as the higher-order relations among universals—can be adequately specified, without reduction, by a set of postulates indicating the roles of these relations within the domain of properties and states of affairs. If he is correct about this, then we have an analysis of the causal relation that can be employed in an account of agent causation. An agent causalist can say that it is the relation thus analyzed that obtains between a person and her action when she acts with free will; it is the very relation that, within the domain of properties and events or states of affairs, occupies the specified role.
>
> Moreover, an account that runs parallel, at a certain level of description, to that suggested for event causation would seem to be available for agent causation. An agent may be held to cause a particular action (more precisely, an event of acting on a certain ordering of reasons) just in case the relation of causation obtains between these two particulars. And an agent can be said to be so related to one of her actions only if these two particulars exemplify certain properties. Perhaps the only agents who cause things are those who have the property of being capable of reflective practical reasoning, and perhaps such an agent directly causes only those events that constitute her acting for reasons. There might, in that case, be a law of nature to the effect that any individual who acts with such a capacity acts with free will.
>
> Natural law, then, may subsume all free action without undermining the freedom with which human beings act. On this sort of account, the agent causation on which free will is said to depend is seen as thoroughly natural. (197–98)

To what extent does this story provide us with an intelligible naturalistic account of agent causation? Is our understanding of agent causation rendered satisfactory by seeing the agent-causal relation as the familiar causal relation with an abnormal relatum, and treating that relation as irreducible? It is worth listing four residual and persistent difficulties of this picture, some of which Clarke is aware of.

(1) Clarke concedes that we have no idea of what it would be like to recognize the agent-causal relation as obtaining: "There is," he says, "no observational evidence that could tell us whether our world is an indeterministic world with agent causation or without it" (199). So our grasp of the truth conditions of agent-causal ascriptions must radically transcend any grasp of what it would be like, even in principle, to recognize agent causation as obtaining. (Some have recommended that at least in the first person we know perfectly well how to recognize agent causation—that the agent-causal relation is manifest to us from the first person in the ordinary context of deliberating and acting on that deliberation. Clarke rejects this in passing, and indeed there are good reasons for rejecting it. In particular, there is excellent reason to think that robust freedom is not always manifested in ordinary deliberation, especially where the deliberation brings to light overwhelming reasons in favor of one action over any other.[65])

(2) Clarke claims that agent causation "would not improve our ability to predict and explain human behavior," adding that, "If prediction and explanation are paradigmatic of scientific understanding, it appears that agent causation neither contributes to nor detracts from such understanding" (199). Why then does Clarke think we ought to believe in robust freedom of an agent-causal sort? Because he thinks that this is presupposed by our ascriptions of moral responsibility to one another. That may well be. But if he is right that the agent-causal picture has no place in the project of gaining a "scientific understanding" of the world, it is not surprising that its status continues to be—and will remain—problematic.

(3) A crucial issue that Clarke does not address—crucial, in particular, given his tentative suggestion that agent causation and event causation alike are grounded in causal laws—concerns the purported connection between agent causation and being able to do otherwise. Let us grant that the causal relation sometimes obtains between agents and events rather than only between events. What would prevent God from issuing the following decree?

> If agent *S* exists at time *t* and has motives m_1, m_2 and m_3, then the causal relation will obtain between *S* and action *a*.

If agent causation is the familiar causal relation, why should God be any less able to decree when and where some agent will stand in the causal relation to actions than He is able to decree when and where events will stand in the causal relation to one another? And if, as seems plausible, God can so decree that this familar causal relation hold, it would hardly appear that "agent causation" entails freedom to do otherwise, and so it would hardly seem that agent causation entails robust freedom. (God's decrees might be more general than the one just expressed. They might be of the form "When agents are in states *x*, *y*, *z*, they will stand in the causal relation to actions of type *F*.")

Note that on the present account, we have not secured freedom even in the weak sense of one's actions being undetermined by event-event causal laws. For, according to the story, there is no reason *a priori* why an action might not be *overdetermined*. If the agent-causal relation is the familiar causal relation, why shouldn't some action be overdetermined by a causally sufficient event and some agent, both standing in the causal relation to that action, just as two causally sufficient events can stand in the causal relation to an action? (It may be that Clarke himself doesn't care so much about being able to do otherwise. In a footnote he writes, "I see no problem in saying that, on the agent causal account, the agent, together with her having certain reasons, jointly deterministically cause her acting on those reasons" (202). But this removes at least one main sort of motivation for agency theory. After all, the incompatibilist will now say, "Suppose one has no choice about agent-causal laws. And suppose an agent has no choice about the reasons in her possession for acting. If some agent-causal laws together with those reasons entail her doing *a*, then whatever the relation between an agent and *a*, it will not be up to that agent whether or not to do *a*." Perhaps Clarke has a more Leibnizian compromise in mind, of the sort we sketched earlier.)

(4) Consider Clarke's suggestion, quoted earlier, that "There might be a law of nature to the effect that any individual who acts with such a capacity acts with free will" (198), where the capacity is that of practical and reflective reasoning. Now on most accounts, including Tooley's, laws of nature are contingent. Where exactly is contingency seen to enter into the present account? The idea might be that it is a contingent fact that when a practical reasoner stands in the causal relation to an action, that constitutes free will. But this will once again leave the concept of free will mysterious. For if it is only nomologically necessary that agent causation be associated with free will, then it is not metaphysically necessary that if the causal relation obtains between an agent and an action, the agent performs the action freely. In that case we should still

want to know what freedom *consists in*. (Does it remain open, for example, that there are possible worlds wherein exist laws of nature to the effect that whenever there is a certain sort of event causation, there is freedom?)

A different way of installing contingency is to make it a contingent fact that when an agent with such-and-such a capacity performs an action, the causal relation obtains between the agent and the action. If this is contingent, is there then a possible world where (say) the causal relation holds between an electron and certain events, or a possible world where the causal relation holds between sleeping people and certain events?

What these questions help to bring out is this: It is simply not clear to human minds why the obtaining of the familiar causal relation between a thing and an event should constitute free will; and this is because the human mind recognizes no combination of ordinary explanatory and causal notions that is *a priori* sufficient for freedom, agency, and moral responsibility. Now, it might be a brute speculation that, necessarily, when the causal relation holds between a thing and an event, the thing freely brings about that event. But this proposal will remain just that—a brute modal speculation. (There are analogies from other domains. We might offer the brute speculation that anything with a certain physical structure will have qualia of a certain sort. But such speculation will be inevitably dissatifying to the human mind because there will be neither empirical nor *a priori* resources for providing a compelling story as to why there couldn't be that physical structure without qualia. The is-ought gap, brought out by Moore's famous Open Question Argument, might also offer a relevant analogy here.)

None of this is meant to show that agent causation is incoherent. It simply makes vivid why agent causation theory will seem unsatisfying. We can, on the one hand, commence by means of a certain reference fixer: there is a special relation *R* holding between things and actions that make things morally responsible for that action. Call the relation agent causation. (This is roughly van Inwagen's strategy.[66]) Beginning in that way, the problem is then to spell out agency theory by integrating *R* with other properties and relations of which we have independent grasp via segments of folk theory or science or *a priori* metaphysics. This has remained very hard to do. It has become increasingly clear that no putative sufficient conditions for free agency constructed out of causal, structural, and explanatory concepts will emerge as *a priori* compelling; and in the absence of compelling *a priori* links, it is hard to see how this gap can be bridged by empirical investigation, seeing how elusive agent

causation is to empirical observation. On the other hand, we can straightaway fill out agency theory by identifying agent causation with some familiar relation or relations from science or folk theory or metaphysics. This is Clarke's strategy. The difficulty then comes in providing oneself with compelling reasons for thinking that the story, thus filled out, tells one what is necessary and sufficient for moral responsibility. And added to each of these respective difficulties is the apparent vacuity of agent causation with respect to recognition conditions and the project of prediction.

So agent causation, if there is such a phenomenon, is not one that we understand well; nor, especially, does it even seem to be a phenomenon that the "natural light" of human beings is capable of understanding very well. Now our present question concerns the compatibility of a robust conception of freedom with a materialist theory of the mind. In attempting to answer this question, one shall, as in any other context, be inevitably hampered by the paucity of our conception of robust freedom. If in the sequel we encounter respects in which a materialist theory of mind is not very well suited to a robust conception of freedom, we cannot pretend that a dualist theory of mind magically renders agent causation fully intelligible or magically provides us with the possibility of a well-developed story about agent causation. The point is worth emphasizing: robust freedom will remain somewhat elusive whether or not we embrace dualism about the mind.

What follows is a sort of progress report on our own efforts to make good on the hunch that agency theory, impoverished as it may be, provides its proponents with considerable reason to resist a materialist theory of mind. It thus provides some reason to question the claim of those like Clarke that agency theory is compatible with a fully naturalistic picture of human agents.

An Outline of Materialist Agency Theory

Let us make a start at fleshing out a materialist agency theory and certain crucial decisions the materialist must make. We'll do this by considering an apparent threat to the supervenience doctrine expressed in Assumption 3. Begin with an inferential principle ("Principle ß") advanced by van Inwagen in his elaboration and defense of a robust conception of freedom:

> Np and N(if p, then q) entails Nq

(where "Np" means "p and no one ever had a choice about p").[67] This principle underwrites the common intuition that freedom is incompatible with determinism. For none of us have a choice about the laws of nature and none of us have a choice about the distant past. Coupled with Principle ß, those platitudes entail that, if determinism is true, none of us have a choice about what we do.[68] We shall be assuming that any robust conception of freedom will subscribe to something like that principle.

It is tempting to suppose that we can use Principle ß to show that the sort of determinism noted in the introduction, at work in the purported supervenience of the mental on the microphysical, can also be shown to be incompatible with free will robustly conceived. Let us pursue this thought, beginning with the supposition that

> If we are ever free, then the property of *an agent's making a free decision at t* supervenes upon the intrinsic microphysical history of an agent up to *t* (from Assumption 3).

And let us add the following thesis:

> T1: No one has a choice about what supervenes on what.

T1 is rather analogous to the view that people have no choice about the laws of nature. But notice that T1 is even more compelling than this latter thesis: since supervenience—as we are understanding the term in its normal usage—is a modal relation of metaphysical necessity, it is very strange to suppose that people have a choice about what metaphysically necessitates what.[69] (Indeed, few think that even God has a choice about that sort of thing.) Meanwhile, anything less full blooded than the claim that the microphysical necessitates the mental would fall short of materialism about the mind. So T1 is extremely compelling.

Suppose now that we add another thesis:

> T2: People don't have a choice about any of their microphysical details.

Surely people don't have a choice about the exact spatial relations among electrons and other microphysical particles making up their microstates. And it seems clear that this thesis, in conjunction with Principle ß and thesis T1, will entail that, if an agent's making a free decision at *t* supervenes upon the intrinsic microphysical history of an agent up to *t*, then we are never free. (Note we are not assuming deterministic physical laws here.)

What is one to say in the light of this argument? One could deny that people are free. Or, one could deny Principle ß and affirm some deflationary view of what it takes to be free. Or, one could give up one or another of Assumption 3, T1, or T2. Since we are interested in the consequences of ascribing robust freedom to people, the first two options—of denying we are ever free and denying Principle ß to adopt a deflationary view—are of no interest in this context. Now T1 seems unchallengeable; and giving up Assumption 3 is tantamount to giving up a materialist theory of mind. Thus, if one is to try to reconcile robust freedom with a materialist theory of mind, one will have to give up T2, affirming that people do indeed have a choice about their microphysical history.

We can go further. Suppose, by hypothesis, that one freely makes a decision at *t* and that Assumption 3 holds. What sort of control of one's microphysical history will this require? It will not be good enough to have control over one's microphysical history after *t*. For if one does not have control over one's microphysical history up to *t*, and if one's decision supervenes on one's microphysical history up to *t*, then Principle ß tells us that one's decision at *t* is not free. Given that we have no choice about the past, we cannot say that by virtue of exercising a choice at *t*, one exercises control over one's microphysical history prior to *t*.[70] So, if we wish to affirm robust freedom and retain Assumption 3, we will be forced to say that if some agent makes a free decision at *t*, then it will be up to that agent what microphysical states he is in at *t*. That is, one will have at a time "top-down" agent-causal control over one's microphysical states at that time.

Suppose we go with this top-down causal picture and see where it takes us. On behalf of the materialist, let us run the following line. Agents are material beings. Facts about whether an agent makes a decision at *t* supervene on the agent's microphysical structure at *t*. Agents are sometimes free. Insofar as an agent is free at *t*, it is up to an agent what microphysical states he is in at *t*. And insofar as "up to" involves agent causation, the agent will thus have the ability to agent cause some of his own microphysical facts, even though these agent-causal facts at *t* supervene on the microphyscial history of the agent up to and including that time.

Is there any logical incompatibility between "agent causation supervenes on microphysics" and "an agent is causally responsible for some of the microphysical facts forming the subvenient base for that very causal responsibility"? There is certainly no inconsistency in the following triad: a family A supervenes on family B, family B doesn't supervene on A, and yet some member of A is causally relevant to some member of B. Supervenience does not straightforwardly entail anything about causality. That is, the modal determination of family A by B doesn't straightforwardly mean that there is no causal determination of some member of B by some member of A. (Here is a simple example, departing perhaps from the sort of supervenience (of nonoverlapping families) most commonly discussed in the literature, but still in keeping with the letter of the requirements for supervenience. Suppose e_1 causes e_2. The family $\{e_2\}$ supervenes on the family $\{e_1, e_2\}$ and not vice versa; and yet a member of the former family causes a member of the latter family.) So there is no cheap and easy way to show that agent *a*'s causing microstate *m* is incompatibile with the supervenience of "*a*'s causing *m*" on a set of microphysical facts that includes *m*.

We can, in this context, get clearer about the sense in which a materialist will affirm that his microphysical states are up to him. Consider the proposition P expressing a precise specification of all one's microphysical states at *t*. One might now think: "It is not up to me whether or not I enjoy that precise arrangement of microphysical states. For I certainly don't have control over, say, the exact location of each of my microparticles. I can now import Principle ß and Assumption 3. Suppose I make a decision at *t*. It is not up to me whether or not P. And it is not up to me whether or not supervenience relations of sort R hold. P and R obtain and together entail my decision at *t*. So my decision at *t* is not up to me." But consider now an analogous argument. Suppose that my arm takes some precise trajectory T. It might be argued that I don't have control over whether my arm takes exactly that trajectory. And my arm moving supervenes on T. Thus, I don't have control over my arm's moving (by Principle ß). But presumably I do; something has gone wrong.

What needs further clarification on behalf of the materialist is the sense of "up to me" that may be reckoned operative here. As van Inwagen implies, so long it is up to me whether my arm moves at all, I do indeed have a choice about that trajectory T.[71] That trajectory is up to me in this sense: T can be avoided by *avoiding arm-moving altogether*. (What is not up to me is whether, *given* that my arm moves, it takes trajectory T.) Analogously, so long as it is in my power to bring it about at *t* that P is *false* (so long as I can avoid decision making altogether at *t*), then even though I don't, as it were, have control over all the exact microphysical details of my state at *t*, I will be such that it is up to me at *t* whether or not P; my decision is up to me.

Let us suppose, then, that we cannot exhibit any logical incompatibility

between "agent causation supervenes on microphysics" and "an agent is causally responsible for microphysical facts at the subvenient base for that very causal responsibility." Having nevertheless gone a short distance toward fleshing out a view on behalf of the materialist agency theorist, it is worth acknowledging some of the decision points that such a theorist will face along the philosophical decision tree. Two such decision points can be framed in the form of questions the materialist must answer.

First: "What do we say about the relation of agent causation to microphysical laws?" We note two main options here.

1. The materialist could say that the normal operation of microphysical laws is disrupted by downward causation. One who embraces this picture might envisage the following sort of scenario:

> The following sequence of microevents occurs: m_1 and m_2 are followed by m_4 and m_5. At the time at which m_4 and m_5 occur, an agent causes m_5. In the world in which this sequence occurs, it is a microphysical law that if m_1 and m_2 occur, then m_3 and m_4 will follow (though the actual sequence in the present instance does not accord with this law). It is a true supervenience principle that if m_4 and m_5 obtain at some time, then an agent causes m_5 at that time.

Is this coherent? That depends in part upon whether there being a law that Fs cause Gs entails there being a universal regularity of Fs being followed by Gs. Clearly, the sort of materialist agency theory just now sketched must bring with it a considerably different picture of laws of nature, a picture allowing for local breakdowns of normal laws for microphysical particles when those particles are caught up in complex systems of a special sort. (Note that this picture might even allow that the laws of nature be deterministic, since freedom is secured by allowing for the possibility that, so to speak, the laws of nature be broken.)

2. Alternatively, the materialist can say that normal microphysical laws are not disrupted, that agent-causal facts are additional causal facts that do not interfere with microphysical laws. (Note that since this picture will not tolerate an agent's breaking the laws of nature, the laws of nature had better be statistical and not deterministic. Were they deterministic, then given their unbreakability and given that we have no choice about the distant past, Principle ß would make trouble for any claim of freedom for our decisions.)

The picture adumbrated by option 1 has been entertained in a recent paper by Tim O'Connor.[72] He notes that if we take option 1 seriously,

we can envisage getting empirical evidence for top-down causality. The proponent of option 2, meanwhile, while having a much harder time telling a story about how to garner scientific evidence for top-down causality, will have the advantage of less conceptual strain.

We turn now to a *second* important decision point: "What exactly should be said about the relation between the willing, the micro-events, and the agent?" We know the materialist will be committed to the supervenience of willing-facts upon microfacts. But there are a number of options consistent with that supervenience thesis:

> (a) We can say that a willing is token identical with a complex microphysical event, and that agent causation relates the agent to that complex event directly.
>
> (b) We can say that a willing is token distinct from micro-events, though supervenient on them, and that the agent causes the willing, which in turn causally influences the micro-level.
>
> (c) We can say that a willing is token distinct from micro-events, though supervenient on them, and that the agent causes some complex micro-event that in turn causes the willing.

It might be thought that there is an epiphemonal alternative, according to which the agent causes a willing (that is token-distinct from microphysical phenomena), but exerts no influence whatever on micro-events (whether directly or indirectly). But this option is not available for anyone who endorses Principle ß. If your willing supervenes on microphenomena then, if Principle ß is correct, it had better be up to you what microphysical states you are in (for otherwise, we can deduce that your willing is unfree).

So the materialist agency theorist is left with (a), (b), and (c) as ways of relating the willing, the agent, and the micro-level. Note that in cases (a) and (b), the willing is proximal to the agent, whereas in (c), it is distal. Surely this third option is less plausible by far. Case (c) is a theory of agency according to which the fundamental agent-causal relation is between an agent and *nonmental events*. We find this picture too far removed from our normal intuitions about agency. In part, at least, one should have thought that deciding is a basic action: one doesn't decide by doing something else.[73] Hereafter, we'll be concerned with the merits of (a) and (b), leaving (c) aside as relatively undeserving of attention. For ease of reference, we can call option (a) Token-Reductive Materialism, and option (b) Token-Emergent Materialism.

That completes enough of our story about what materialist agency

theory might look like. We now switch to a more polemical mode: our aim is to indicate why on balance an immaterialist agency theory appears more rational than a materialist one. In this connection, it is (again) no use pointing to costs that must be paid by any agency theory, whether of a materialist or immaterialist stripe. As we have said, any agency theory is likely to remain elusive with regard to details. Similarly, any agency theory may well have to reject what Lewis dubs "the explanatory adequacy of physics," according to which "there is some unified body of scientific theories, of the sort we now accept, which together provide a true and exhaustive account of all physical phenomena."[74] Shy of projecting that agency theory is to become part of natural science (a dim prospect, we conjecture), the agency theorist cannot accept the complete explanatory adequacy of natural science. (And at any rate, it certainly does not seem that an explanatory theory in which agency theory is central would be a scientific theory "of the sort we now accept."[75]) Considerations of this latter sort, if correct, provide no reason for preferring a nonmaterialist agency theory over a materialist agency theory. What we need to do is identify the special costs incurred by materialist agency theory that are not incurred by a dualist agency theory.

We shall discuss, in turn, the two central ideas of materialist agency theory—first, the supervenience thesis adumbrated by Assumption 3, and second the thesis that the agent is a material being.[76]

Costs of Combining the Supervenience Thesis with Agency Theory

We begin with two general difficulties for the materialist agency theory as sketched. A first point to acknowledge is that the materialist agency theorist offers a picture that is at best strained, in the following respect. We are invited, on the one hand, to imagine that certain microphysical P-facts metaphysically determine (thanks to supervenience) distinctively mental agent-theoretic M-facts, while yet (on the other hand) admitting that certain agent-theoretic M-facts are causally relevant to some of those very microphysical P-facts upon which they supervene. By our lights, this combination of theses will strike any reader who fully absorbs them as intuitively odd, odd in *something like* the way that talk of self-causation is odd. It would be nice to articulate what is odd about this combination in terms of one or more intuitively compelling principles that must be violated by this version of agency theory. The best we can

offer here is the following principle (we invite readers to see if they can do better):

> No state of affairs x can metaphysically necessitate some state of affairs y if y is causally relevant to the obtaining of x.[77]

Suppose that facts about agency supervene on microphysical facts and not vice versa. It will then be hard to deny that there is a minimally sufficient supervenience base that necessitates the agent-causal facts one level up. By the above principle, there can be no agent-causal explanation for the obtaining of that subvenient base. So if the principle is correct, materialist agency theory cannot be sustained, since it precisely does claim that the microphysical base necessitates agent-causal states of affairs that causally explain that base itself.

Is the above principle true? While it falls short of anything like self-evident, we find it compelling. By our lights, the fact that materialist agency theory needs to reject the principle must be reckoned a cost of the theory.

A second cost to acknowledge is the following. Metaphysicians often find attractive the idea that supervenient facts are nothing over and above the facts that they supervene upon. In David Armstrong's characterization (in conversation), supervenient entities are "an ontological free lunch." Now, clearly, our materialist agency theorist cannot buy into this picture. On his account, agent causation is one of the "joints of nature," hardly assimilable to some purely microphysical relation or some gerrymandered disjunction of them. Nor could the picture of downward causation be readily sustained if the M-facts of agency were "nothing over and above" the P-facts of microphysics. What the agency theorist needs is some sort of "emergentism" according to which, though agent-theoretic M-facts supervene on the microlevel, they are in some important sense "something over and above" the microphysical P-facts. He will thus postulate some sort of necessary connections (or, in an alternative lingo, "internal relations") between distinct existences—between facts that are both conceptually and genuinely distinct.[78]

Now clearly this internal relation is nothing like an analytic one. It is surely *conceivable* that there be a possible world whose microphysical structure is just like this one, with the same statistical laws of nature in operation and where there is no agent-causation that brings about some arm movement that at this world is freely brought about by a person in an agent-causal manner.[79] At that world, the arm movement is an undetermined event that is not caught up in that agent-causal relation.

While conceivable, the materialist agency theorist cannot allow that this scenario is possible. Nor can this internal relation be merely nomological, since nomological connections are weaker than necessary connections. In short, it looks as if materialist agency theory presents us with a particularly puzzling brute necessary connection between distinct existences.

If it is an advantage of a theory to minimize obscure necessary connections between distinct existences, then it will be an advantage for the proponent of robust freedom to avoid materialism about the mind. Of course, it would be no improvement to say that agents are immaterial and then go on to claim that agent-causal facts supervene on quasi-structural ectoplasmic properties. One would be left once again with obscure necessary connections between distinct existences. The difficulty can be avoided only by an agency theory that makes agent-causal facts nonsupervenient.

So much for general costs of the supervenience thesis. As we said earlier, it can be combined with token monism or token dualism. We outline below the costs of each particular combination.

A Special Cost of Token-reductive Materialism

Suppose some neural event is a willing. Is it essentially a willing? Many token identity theorists talk as if neural events only enjoy mental descriptions contingently. If that is so, a special worry seems to confront the materialist who believes in robust freedom. For it is arguable that events are capable of entering into the fundamental causal relations by virtue of their intrinsic, essential natures and not their accidental, purely circumstantial properties.[80] If agent causation is a fundamental causal relation, and being a willing is a merely contingent feature of an event, then it does not seem that events enter into the agent-causal relation by virtue of their being willings. But then it is hard to see why an event that is not a willing couldn't enter into that relation. This conflicts with the fairly compelling intuition that agent causation, if there is such a thing, is restricted to willings; it is by virtue of their being *willings* that those events are fit relata for agent causation.

One might insist that being a willing is a contingent, relational (circumstantial) property of an event, and preserve the claim that *necessarily agent causation produces willings* by reckoning it analytic: "being a willing," on this account, would just be understood to mean "being an event that is the proximal effect of agent causation." But on this proposal, one does not genuinely *explain* the suitability of an event for being the proximal effect of agent causation by saying that it was a willing.

Consider any event kind K and assume events of that kind get caught up in fundamental causal transactions. There will then be truths to the effect that events of kind K get caught up in certain sorts of fundamental causal transactions but not others. Call truths of this sort the "fundamental causal truths about Ks." We're supposing that there are fundamental causal truths about willings. If we invoke the causal-explanatory premise that a perspicuous explanation of fundamental causal truths about Ks will proceed by invoking intrinsic, nonrelational properties of Ks, then on our materialist's relational treatment of being a willing, we arrive at the conclusion that one cannot explain the fundamental causal truths about willings by appealing, in part, to their status as willings. But this conflicts with our initial conjecture that it is by virtue of their status as willings that certain events are fit relata for agent causation.

Whether or not the property of being a willing is essential to an event, it is worth underscoring the fact that nearly all philosophers of a naturalistic bent assume that the preferred description for a mental event vis-à-vis the project of deep explanation is as a neural event of a certain sort. Thus, for example, Donald Davidson's anomalous monism insists that mental events need to be redescribed neurally if they are going to become fully intelligible *qua* events in the natural causal order.[81] There is good reason for this. It is very plausible to think that the natural kinds to which neural events belong are neural kinds, and moreover that deep explanation of events will proceed by considering events *qua* the natural kinds to which they belong. But if agent causation holds between agents and neural events, we shall have something like the flip side of anomalous monism. Agency theorists of a materialistic bent will have to maintain that the special agent-causal relation holds between material things and a particular sort of neural event when those material things and neural events are redescribed in terms of the categories of moral psychology. (And, if by the lights of token identity theory the neural events that are in fact willings don't form their own natural kind, it would seem odd at best to postulate a unique (agent-) causal relation that's restricted to an event-type that is not a natural kind among events.)

A Special Cost of Token-emergent Materialism

Token-emergent materialism claims that there is a mental event—a willing—that is distinct from any microphysical event and that an agent influences its microphysical states by agent-causing a willing, which in turn causally influences how microphysical states are distributed. A cost

here, as we see it, is the violation of a moderately compelling Humean thesis about causation. The thesis is the following:

> No causal relation obtaining between a member of a family of events A and some member of a distinct family of events B is necessitated by the existence of A.

Token-emergent materialism violates this principle in the following way. It tells us that some family of microphysical events is sufficient for the existence of some willing distinct from the microphysical events and sufficient for that willing's being causally related to one or more of those microphysical events. So it appears that in order to combine supervenience with downward causation of microphysical phenomena by willings, a staunchly anti-Humean position on causation must be defended. Again, while we have identified this as indeed a cost, nevertheless—like all those commitments of the supervenience thesis that we have identified as costs—it is imaginable that someone would deny this, or believe it a cost worth paying.

Costs of the Thesis That Agents Are Material

We end with several reasons for worrying about treating the agent that enters into agent causation as a material object. These will provide at least some reasons for an agency theorist's believing that the agent is immaterial.

In answer to the question "What physical object is the agent?," we have been thinking that persons, according to a materialist agency theorist, are physical organisms, and that such organisms are the agent-causes of their decisions. Earlier we took seriously the idea that the fundamental causal powers of an event are determined by its intrinsic, essential properties.[82] Consider now the following, closely related, thesis—which also strikes us as true:

> The fundamental causal powers of a thing aren't determined wholly or partly by properties that are extrinsic to it.

Applying this principle now to agency theory, we can infer that the causal power of a thing to produce decisions doesn't depend upon its environment; if two things are intrinsically qualitatively identical, then either none or both are agent-causes of decisions. But this truth seems

jeopardized by materialist agency theory. In the actual world, Jan agent-causes his decision to stop daydreaming about climbing big mountains and to work on philosophy instead. Suppose the agency theorist identifies Jan with the salient physical organism. Consider now world W where the decision is made by a brain in a vat instrinsically identical with the one Jan enjoys in this world, surrounded by 145 pounds of organism-stuff. In the actual world, the brain does not agent-cause the decision, but in the brain-in-vat world, it does. We have violated the principle. Thus the materialist agency theorist will have to either give up the principle or claim that even in the actual world, agents are brains and not physical organisms.

The second option strikes us as *prima facie* more plausible, though it is still troubling: there is an undeniable strain to claiming that strictly speaking moral predicates of praise and blame apply to brains. Moreover, further trouble looms ahead. Shall we say all or some proper part of the brain is the agent? Suppose that we are materialists and on some occasion identify the right hemisphere of an individual as most intimately involved in some decision that is agent-caused. Consider now a right-hemisphere-in-vat world. . . .

Even ignoring the compelling nature of the above principle, this latest difficulty highlights a certain awkwardness in the task of saying what physical object does the agent-causing. The organism? The *whole* organism, feet and arms and all, or some proper part of it? The brain, perhaps; or perhaps one hemisphere, or some smaller part. Sometimes this hemisphere, sometimes that. The whole business looks to emerge as arbitrary, in a way that no facts-of-the-matter are.[83]

Few thoroughgoing naturalists would shy away from allowing that, at the end of the day, there must be some principled way of marking off what physical object as a matter of fact does the deciding. This said, it is worth noting that one of the most attractive conceptions of material things is one according to which the really fundamental relations holding between material things hold between the microphysical particles, at the deepest level of physical theory. Whatever agents may be, it will strike many of us as both odd and a surprising concession to have very large material beings enter into the fundamental explanation of many microphysical events.

A while back, H. Feigl and then J. J. C. Smart worried about nomological danglers: they could not bring themselves to believe that physics had all its fundamental laws holding among tiny particles save for a few special laws holding between large complexes of millions of particles and conscious states. Their conclusion, or anyway Smart's, was that "man is

a vast arrangment of physical particles and that there are not, over and above this, sensations or states of consciousness." If we are to accede in the claim that man is a vast arrangement of particles, there is considerable pressure to hold that the microlevel is metaphysically prior in order of explanation. We want to leave fundamental physical explanation all at the level of the microparticles, where deep physical explanations of physical things ultimately belong.

Is there any advantage for dualistic over materialistic agency theory here? Well, dualism allows one to stick with the idea that the fundamental material processes are at the microphysical level. That is to say, all fundamental transactions between wholly material things are at the microphysical level. It also allows one to stick with the intuition that the fundamental explanations for such transactions proceed by appeal to laws of nature. We've already seen how difficult it is for robust freedom to be grounded by laws of nature. If one wants robust freedom and also thinks of purely natural processes as grounded by laws of nature, then one had better place agents outside the natural world. In short, it seems to us that dualism affords the proponent of robust freedom the prospect of a more orderly and unified conception of the material world, of material things. Relatedly, the immaterialist is not forced to graft a radical difference in kind onto what appears to be a mere difference in degree. If it is acknowledged that agent causing is, as it seems, a radically unique kind of making or bringing about relation, the materialist agency theorist must acknowledge that there is such a relation out in the world that crops up in certain physical organisms. These organisms differ from other organisms, trailing off in degree from chimps to bunnies to lizards and on down. The materialist must say that the difference in degree between the organisms that are our bodies and those of, say, monkeys, produces a radical difference in kind. It must surely be a point in favor of the immaterialist that a difference in kind, when it comes to unique causation and agency (and the moral responsibility brought with it) is marked by an ontological difference between things capable of entering into such causation and things incapable of entering into it.

Finally, to return to an earlier point, agent causation is a genuinely mysterious sort of phenomenon. If it occurs at all, one might reasonably expect it to emanate from a very mysterious sort of thing. And one might reckon organisms or brains scarcely mysterious enough: souls, substantial forms, and entelechies seem eminently suited to that role.[84]

Conclusion

We have found nothing outright incoherent about materialist agency theory. If the materialist accepts, for example, that for certain deep explanatory purposes, the flip side of anomalous monism should be accepted, doesn't mind saying that there are deep explanations of microphysical events that are provided by certain hulking large material objects, isn't especially attracted to a simple and unified conception of material beings, doesn't mind saying that the agent-causal relation is constituted in part by facts extrinsic to the relata, and so on, then he will find the marriage of agency theory to materialism relatively untroubling—certainly little more troubling than agency theory is in its own right. What we have aimed to do is alert agency theorists to the costs of embracing materialism, and leave them to decide whether these costs—ones rather less heavy than embracing contradictions or denying positively self-evident principles—are prices worth paying.[85] While we shan't pretend that the cumulative weight of the points we've raised against materialist agency theory are overwhelming, perhaps it will at least be clearer why we found ourselves with the belief that there are connections of interest and importance between freedom and materialism.[86]

Chapter 3

56. A complication: As Daniel Nolan pointed out to us, it is arguable that microphysical laws at a world do not supervene on the microphysical states at a world; and a materialist might say that mental properties are determined by microphysical states plus laws relating microphysical properties to each other (call them "micro-micro laws"). Fair enough. We invite materialists of that stripe to extend "microphysical states" to include micro-micro laws.

57. See Hilary Putnam's "The Meaning of 'Meaning,' " in his *Mind, Language and Reality* (Cambridge: Cambridge University Press, 1975) and Tyler Burge's "Individualism and the Mental," *Midwest Studies in Philosophy IV* (Minneapolis: University of Minnesota Press, 1979). Standard Twin Earth type cases, adduced to show that an individual's mental states do not supervene on the microphysical states of that individual, are cases where duplicate individuals are in different physical environments: such cases do not, then, have any force against the thesis that the distribution of mental properties supervenes on the global distribution of microphysical properties. The volume, edited by P. Pettit and J. McDowell, *Subject, Thought and Context* (Oxford: Oxford University Press, 1986) is devoted to this topic of externalism and mental content.

58. As Frank Jackson notes in "Armchair Metaphysics," J. O'Leary-Hawthorne and M. Michael, eds., *Philosophy in Mind* (Dordrecht: Kluwer, 1994), "Physicalists are typically happy to grant that there is a possible world exactly like ours but which contains in addition a lot of mental life sustained in non-material stuff" (28), and thus, when they say that the distribution of mental properties supervenes on the microphysical properties, there must be some tacit restriction of the worlds considered. Jackson's suggested reformulation on behalf of the materialist: "Any world which is a minimal physical duplicate of our world is a duplicate simpliciter of our world" (28)—though he admits to there being no rigorous, noncircular definition of "minimal physical duplicate."

59. Again, angels share all of their microphysical properties (none), but have different mental properties. Thus the qualifier about immaterial beings.

60. Though we should pause long enough to acknowledge "nonstandard" breeds of compatibilism denying that freedom simply amounts to actions being appropriately causally connected to past beliefs and desires, even if those actions are contingent upon circumstances beyond the control of the agent. So-called multiple-pasts or altered-past compatibilism suggests that there is a relevant sense in which we can "affect" the past (that is, roughly: there are actions such that, if—contrary to fact—I were to perform them the past would have been different). So-called local-miracle or altered-law compatibilism suggests that there is a relevant sense in which we can have "power over" laws (roughly: there are actions such that, if—contrary to fact—I were to perform them, a law-breaking event

would have occurred). To the extent that they aim to capture some robust notion of power to do otherwise, these forms of compatibilism are not deflationary. (Whether the operative senses in which we can "affect" and "control" the past and laws of nature are relevant to ascriptions of freedom and moral responsibility is another question.)

61. See "How to Define Theoretical Terms," *Philosophical Papers, Volume I* (Oxford: Oxford University Press, 1983), 78–95, esp. 95.

62. As Michael Smith puts it (though he is not an agency theorist), freedom requires that we be the "authors" of our actions. See his "Freedom, Reason and the Analysis of Value," forthcoming.

63. The general picture is there in the earlier sections of the *Discourse on Metaphysics*, its connection with freedom coming in later sections, where Leibniz will say, for example, that "every substance has a perfect spontaneity, which becomes freedom in intelligent substances," and that all its actions are "the result of its own concept or being" (DM §32 at G, IV, 458: L, 324). Thus could Leibniz agree with Spinoza at least this far: a free agent is determined only by "the necessity of its own nature" (G, I, 150: L, 197).

64. *Noûs* 27 (1993): 191–203. Page references to this article will be included in the text.

65. See, for example, Peter van Inwagen, "When is the Will Free?," *Philosophical Perspectives* 3 (Atascedero, Calif.: Ridgeview Press, 1989), 399–422.

66. Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983).

67. Van Inwagen's Principle ß is presented and discussed in *An Essay on Free Will*, 93–105. As worded, the principle needs (modest) repair; see, for example, David Widerker, "On an Argument for Incompatibilism," *Analysis* 47 (1987), 37–41, Timothy O'Connor, "On the Transfer of Necessity," *Noûs* 27 (1993), 204–18, and especially a recent discussion by Thomas McKay and David Johnson, "A Reconsideration of an Argument Against Compatibilism" (unpublished). The required repair does not substantially affect the content of this paper: assuming that one can proceed on the basis of a false principle that approximates a true one if the difference is irrelevant to the current line of inquiry, we continue with ß as formulated.

68. Of course, the altered-past and altered-law compatibilists (see note 60) will deny these platitudes, and so will deny their role with Principle ß in showing that determinism is incompatible with robust freedom.

69. Thus the strategy of altered-law compatibilism—of exploiting the fact that laws of nature are contingent in arguing for the power to do otherwise—is unavailable here, just as it is not available in arguing for the compatibility of human free will and divine omniscience (since God's existence and omniscience are reckoned necessary).

70. It is here that altered-law compatibilism will propose a noncausal sense in which the past is in our control, or up to us. The moral relevance of this innocuous sense of "up to me" is less obvious than that of the proposal to follow, below (see also note 71).

71. See *An Essay on Free Will*, 233–34. The thought that we don't exercise a high level of fine-grained control leads us to think that, say, I don't have a choice about whether I toss a six with the die, or that I don't have a choice about the exact trajectory T of my arm. Van Inwagen's point is that this latches onto the wrong sense of "up to me": there is a fair sense of "up to me," relevant to freedom and moral responsibility, that is at work in claiming that the particular trajectory T is up to me insofar as I can avoid arm-moving itself. Or, in van Inwagen's words, "Strictly speaking, Alfred does have a choice about whether he throws a six, at least provided he has a choice about whether he plays dice. He can avoid throwing a six by avoiding playing dice."

72. "Emergent Properties," *American Philosophical Quarterly* 31 (1994): 94–104.

73. See, for example, Hugh McCann, "Volition and Basic Action," *Philosophical Review* 82 (1974): 451–73.

74. "An Argument for the Identity Theory," *Philosophical Papers, Volume I*, 105.

75. There is one other way to try to blend agency theory with a claim of explanatory adequacy of science. Recall that Clarke's view is that agent causation contributes *nothing* by way of explanation. Perhaps, then, in response, the agent-causation theorist should deny that his theory provides explanations of microphysical happenings, thus leaving open the possibility that some body of scientific theories is exhaustive as far as explanation of physical phenomena goes. But that seems quite wrong. On a promising first pass, an explanation is an answer to a "why" question; and agency theory purports to provide us with answers to why questions. "Why was the agent in such-and-such microphysical states at that time?" "Because of such-and-such exercise of free will." If it is up to me whether I decide to do *a* at *t*, and if that will ensure that electron *e* is not at spacetime location *pt*, then it would appear to be appropriate, supposing I do refrain from deciding to do *a*, to explain the fact that electron *e* is not at *pt* by appealing to facts about agent causation.

76. Note that neither of these strictly entails the other. A sort of parallelism wherein the agent is immaterial is consistent with the supervenience thesis; meanwhile, an agent with wholly material parts might conceivably enjoy states that do not supervene on microphysical ones.

77. We shan't go into the metaphysics of states of affairs, here. In terms roughly hewn, they are the "truthmakers" (to borrow a term from D. M. Armstrong) for true propositions, and have their temporal specifications built in. The principle is thus meant to be, so to say, "at a time." We're grateful to Rod Bertolet for pointing out two flaws with an earlier formulation (his concern involved the conceivability of mutual causation, as illustrated by cases of symbiosis).

78. Note that the token-reductive materialist cannot comfortably say that the causal relation between the agent and the material event that is the willing is analyzable into groundfloor physical relations, and so even he will be committed to a sort of emergentism.

79. Conceivable at least in the sense that no semantic/analytic rule is violated by its postulation.

80. At any rate, we regard this as a compelling conjecture. See Sydney Shoemaker's discussion of nonrelational properties and causal powers of events in "Causality and Properties," ed. Peter van Inwagen, *Time and Cause* (Dordrecht: D. Reidel, 1980), 109–35; and consider also that if it is the purely accidental, external circumstances that place events into their causal nexus, events would emerge as the bare loci of change, with no subsequent events being excluded (save the logically impossible), since nothing intrinsic and essential to any event—nothing by its nature—precludes its entering into some causal relations and requires its entering into others. That, presumably, is not what we take genuine caused changes to amount to.

81. See, for example, "Mental Events," in *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), 207–27.

82. The emphasis on "fundamental causal powers" is important and, we hope, intuitive enough. There is certainly a sense in which powers are context-dependent. A being on one planet may be able to see and a duplicate on another planet be unable to see. Why? Because of differences in light between one planet and another. We owe the example to Jonathan Bennett.

83. Of course, we can't assume that the boundaries of the agent will be unproblematic on every nonphysicalist ontology—though, it is a commonplace of traditional immaterialist accounts that minds or souls are simple.

84. We owe this point to Ted Warfield (in conversation).

85. Finally, this, for theists who may find themselves sympathetic with the Cartesian hunch that "it is above all in virtue of the will that I understand myself to bear in some way the image and likeness of God" (in *Philosophical Writings of Descartes*, eds. John Cottingham, Robert Stoothoff and Dugald Murdoch [Cambridge: Cambridge University Press, 1984], Volume II, 40), or with Leibniz's claim that "the root of human freedom is in the image of God in man" (Gr 300). We have volition as part of our nature. And so does God (since He loves us and is worthy of our praise, both of which imply freedom). The power to will marks out a joint in the world; it is, for lack of a better pair of words, a natural kind. It seems strange to claim that a natural kind supervenes on a certain set of properties in me, to which kind God also belongs but without its supervening on anything. It is strange to think that being a horse (consider Trigger) is a natural kind, but that horses might well have been made of nuts and bolts; it is strange to think that being an agent (consider God) is a natural kind, but that an agent might have been made of, well, nuts and bolts or flesh and bone. Of course, if willing were a boring high level property of functional architecture, then none of this would make much sense. But if, as the robust freedom theorist is wont to say, willing is one of the fundamental making relationships in the order of things, then the above consideration will surely have some weight.

86. A distant (historically thicker, philosophically thinner) ancestor of this chapter was presented to the Society of Christian Philosophers at the 1994 Pacific

Division APA meeting in Los Angeles. We are indebted to Rod Bertolet, Hud Hudson, Daniel Nolan, Tim O'Connor, Ted Warfield and Dean Zimmerman for helpful discussion and comments on an earlier draft. We're especially fortunate to have profited over the years from the good counsel and philosophical influence