

## **Agent Causation**

Timothy O'Connor

in T. O'Connor, ed., *Agents, Causes, and Events: Essays on Indeterminism and Free Will*  
(New York: Oxford University Press, 1995), 173-200.

### **I Introduction**

A natural way of characterizing our typical experience of making decisions and acting upon them - one that would, I think, gain widespread assent - goes something like this: When I decide, say, to go for a walk on a cool autumn evening, I am conscious of various factors at work (some consciously articulated, some not) motivating me either to do so or to do something else instead. And there are some courses of action which, while it is conceivable that I might choose to follow them, are such that they do not represent 'genuine' possibilities for me at that time, given my current mood, particular desires and beliefs, and, in some cases, long-standing intentions of a general sort. But within the framework of possibilities (and perhaps even relative likelihoods) that these present conative and cognitive factors set, it seems for all the world to be up to me to decide which particular action I will undertake. The decision I make is no mere vector sum of internal and external forces acting upon me during the process of deliberation (if, indeed, I deliberate at all). Rather, I bring it about - directly, you might say - in response to the various considerations: I am the source of my own activity, not merely in a relative sense as the most proximate and salient locus of an unbroken chain of causal transactions leading up to this event, but fundamentally, in a way not prefigured by what has gone before. Or, again, so it seems.

But a thesis that enjoys unusual consensus among contemporary philosophers is that this pretheoretic conception is not at all like the way things really are with respect to ordinary human activity. Indeed, most would claim that any attempt to theoretically articulate this commonsense picture of agency will inevitably be incoherent or, at best, irremediable mysterious. However, arguments on behalf of this thesis are not nearly as strong as the confidence with which it is generally held. This observation, together with an examination of the nature of such arguments, leads me to suspect that many philosophers are deeply in the grip of a certain broad picture of

the physical world, one which has come to seem overwhelmingly obvious to them, despite the fact that it rests, so far as I can see, on certain empirical assumptions that are as yet unsubstantiated. I will address this intoxicating picture below, though I fear that my rhetorical skills are not up to the task of breaking the grip it has on some.

In what follows, I will contend that the commonsense view of ourselves as fundamental causal agents - for which some have used the term "unmoved movers" but which I think might more accurately be expressed as "not wholly moved movers" - is theoretically understandable, internally consistent, and consistent with what we have thus far come to know about the nature and workings of the natural world. In the section that follows, I try to show how the concept of 'agent' causation can be understood as a distinct species (from 'event' causation) of the primitive idea, which I'll term "causal production", underlying realist or non-Humean conceptions of event causation. In section III, I respond to a number of contemporary objections to the theory of agent causation. Sections IV-V are devoted to showing that the theory is compatible with ordinary reasons explanations of action, which then places me in a position to respond, in the final section, to the contention that we could never know, in principle, whether the agency theory actually describes a significant portion of human activity.

Let me be clear from the outset about two tasks that I do not propose to undertake here. First, I will in no way attempt to argue or adduce evidence for the claim that the theory described actually applies to human action. (I will, however, briefly suggest what sort of considerations could count as evidence in favor of its applicability.) Nor will I attempt to address the epistemological question of whether it is reasonable to suppose, in the absence of strong, directly-confirming evidence, that the agency theory gives a correct schematic account of (a significant portion of) human activity, though I am inclined to answer this in the affirmative. What follows is strictly an essay in "descriptive metaphysics", charting the internal relationship among concepts in what I believe to be part of the commonsense picture of the world.

## II Event Causation and Agent Causation

I begin with a strong, highly controversial assumption about the general concept of causality. This assumption is that the core element of the concept is a primitive notion of the ‘production’ or ‘bringing about’ of an effect. This entails the negative thesis that a satisfactory reductive analysis of causality along Humean lines (in any of its versions) cannot be given. It should be readily apparent that if, contrary to this anti-Humean assumption, a satisfactory reductive analysis of causality can be given, the agency theorist's project of defending a variant species of causality immediately collapses into incoherence. For such reductive analyses are either committed to a general connection between certain types of causes and effects or equate causation with a form of counterfactual dependence. Neither approach is consistent with the agency theorist's claim that a causal relation can obtain between an agent and some event internal to himself, since his understanding of this is such as not to imply that the sort of event effected on that occasion will or would always (or generally) be produced given relevantly similar internal and external circumstances.

Acceptance of this assumption naturally (though not inevitably) points one in the direction of some sort of ‘necessitarian’ account of event causation. It is debatable, of course, whether the necessity in question is to be identified with broadly logical (or metaphysical) necessity or is rather to be thought of as a special, contingent form. Though the accounts I rely on in sketching a broadly necessitarian view take the former route, all that I want to assume here is that there is some form or other of objective necessity attaching to event-causal relations, as it is quite compatible with my purposes that this be held to be a primitive, contingent variety.<sup>1</sup> (Let me forestall confusion by emphasizing at the outset that, in drawing upon the necessitarian view of event causation in order to explicate the notion of agent causation, I am not suggesting that there is anything analogous to a necessary connection between prior circumstances and agent-caused events. Indeed, I will argue below that this is impossible. Rather, as will become clear shortly, the sole aspect of the necessitarian view that I carry over to agent causation is the necessary connection between an object's instantiating a certain set of properties and its possession of a

causal power or powers. The two sorts of causation differ sharply, however, in terms of the manner in which causal powers are exercised: of necessity, when the object is placed in the appropriate circumstances, for event causation; under the voluntary, unnecessitated control of the agent, for agent causation.)

A recent elucidation of a necessitarian approach to causality is found in Harré and Madden (1975)<sup>2</sup>. The central notion in their theory is that of the "powerful particular". When placed in the appropriate circumstances, an object manifests its inherent causal powers in observable effects. The particular powers had by a given object have their basis in its underlying nature - its chemical, physical, or genetic constitution and structure. Events figure in the causal relation in virtue of "stimulat[ing] a suitable generative mechanism to action, or [clearing away] impediments to the activity of a powerful particular already in a state of readiness to act."(p.5) An example of the first sort of causal event is the detonation of a stick of dynamite. The other sort - the removal of an impediment to action - is exemplified by the removal of the air from an underwater cylinder, thereby enabling the body of water to exercise its power to crush the object. Certain effects are 'characteristic' of objects in the appropriate circumstances in a strong sense - "given the specification of the causal powers of the things and substances of the world, the denial of statements describing these effects of those powers, when the environment allows them to be exercised, would be inconsistent with the nature of those things" (p.15). "While natures are preserved, the world must go on in its usual way", although "[n]ecessity might, and probably, does, hold in some cases only between the productive circumstances and a certain distribution of possible outcomes or productions" (p.153).

Now it is natural to link the causal powers an object possesses at a given time directly to its properties. Shoemaker (1984)<sup>3</sup> provides a helpful explication of this idea. First we are told that the possession of a causal power by an object is to be thought of as its being the case that "its presence in circumstances of a particular sort will [of necessity] have certain effects" (p.211). Properties figure into the picture in the following way:

Just as powers can be thought of as functions from circumstances to causal effects, so the properties on which powers depend can be thought of as functions from properties to powers (or, better, as functions from sets of properties to sets of powers). One might even say that properties are second-order powers; they are powers to produce first-order powers (powers to produce certain sorts of events) if combined with certain other properties.

(p.212)

This implies that the relationship between an object's properties and its causal powers is a logically necessary one: "what makes a property the property it is, what determines its identity, is its potential for contributing to the causal powers of the things that have it." (p.212)

If one wishes to hold, by contrast, that causal necessity is logically contingent, then one may say that properties are contingently associated with such functions from properties to powers, rather than being identified with (or logically connected to) them. And another possible wrinkle on the broad position, as Shoemaker notes, is to allow that

the [causal] laws. . . may be statistical, the powers to which the properties contribute, may, accordingly, be statistical tendencies or propensities, and the causation may be nonnecessitating.

(p.232)

With this thumbnail sketch of a standard necessitarian account of event causation before us, I will now turn to the central task of showing how the notion of agent causation may be seen as a distinct species or embodiment of the basic, primitive notion of causal production. The core idea is quite simple: First of all, according to my preferred understanding of the agency theory, wherever the agent-causal relation obtains, the agent bears a property or set of properties that is volition-enabling (i.e., in virtue of this property, the agent has a type of causal power which, in accordance with tradition, we may term "active power"). In this way, then, claims of the form "agent A caused event e" also satisfy a weak version of Davidson's Humean dictum that "causal

statements are implicitly general": such assertions may be thought to imply that a similarly-situated agent (i.e. such that the relevant internal and external properties are instantiated) will always have it directly within his power to cause an event of the e-type.

Thus, the agency theory (as I interpret it) affirms the completely general claim (i.e., one applicable to both of the basic sorts of causation) that objects have causal powers in virtue of their properties, so that objects sharing the same properties share the same causal capacities, but it denies that all such causal powers may be thought of as (or as being intimately associated with) simple "functions from circumstances to effects" (as Shoemaker puts it). For it maintains that some properties contribute to the causal powers of the objects that bear them in a very different way from the event-causal paradigm, in which an object's possession of property P in circumstance C necessitates or makes probable a certain effect. On this alternative picture, a property of the right sort can (in conjunction with appropriate circumstances) make possible the direct, purposive bringing about of an effect by the agent who bears it.

Such a property thus plays a different functional role in the associated causal process. It gives rise to a fundamentally different type of causal power - one that in suitable circumstances is exercised at will by the agent, rather than of necessity, as with objects that are not partly self-determining agents.

To repeat, then, the fundamental tenet of the agency theory may be taken to be the claim that there are two basic sorts of (causal) properties, one of which applies uniquely to intelligent, purposive agents.<sup>4</sup> The thesis that there are two fundamental sorts of causation is a consequence of the thesis concerning types of properties.

Now some may be willing to grant the basic internal coherence of this alternative paradigm, but will maintain that special assumptions would have to be made concerning the nature of the agent in whom such a property were instantiated, assumptions that are not plausible. The most common thought here is that it presuppose some form of substance dualism.

Let us consider, therefore, the compatibility of the agency theory with the view that the only substances to be found in the natural world are material substances. (I intend this to be

noncommittal on the question of whether certain material substances such as living human brains can have irreducibly mental (i.e., nonphysical) properties.) A human agent, in particular, is a wholly biological organism, whose macro-properties are either constituted by or supervene upon the properties of certain elementary physical particles, organized into complex sub-systems at a number of levels. Now some philosophers, it seems, are convinced that this basic picture inexorably leads to the following:

Since all of the surface features of the world are entirely caused by and realised in systems of micro-elements, the behavior of micro-elements is sufficient to determine everything that happens. Such a 'bottom up' picture of the world allows for top-down causation (our minds, for example, can affect our bodies). But top-down causation only works because the top level is already caused by and realized in the bottom levels.<sup>5</sup>

But why does the author (John Searle) consider it an assured result that this "bottom up" picture is applicable to everything that happens in nature? Certain passages in the text from which this quotation is taken seem to suggest that it simply follows from the view that nature consists of material substances built up out of elementary particles, while others may be read as claiming that there are strongly-confirming empirical grounds.

Surely the former reason is without merit. How can we deduce a priori that the organization of matter into certain highly complex systems will never result in novel emergent properties - either properties that themselves exert (in certain circumstances) an irreducibly "downward" form of causal influence, or ones which enable the objects that bear them to do so "at will"? Thomas Reid saw this point clearly. Although he was a substance dualist who thought that no purely material substances are capable of thought, he considered the implications for material agency if he were mistaken in this assumption:

[But if matter] require only a certain configuration to make it think rationally, it will be impossible to show any good reason why the same configuration may not make it act rationally and freely. . . .Those. . .who reason justly from this system of materialism, will easily perceive, that the doctrine of necessity<sup>6</sup> is so far from being a direct inference, that it can receive no support from it.<sup>7</sup>

Unfortunately, I haven't the space here to explore at any length the concept of an emergent property on which I'm relying. Suffice it to say that an emergent property is a macro-property which supervenes upon the properties of an object's micro-structure, but whose role in the causal processes involving that object are not reducible to those of the micro-properties.<sup>8</sup> I'm inclined to think that any tendency to suppose that the emergence of macro-determinative properties in material substances is strictly inconceivable must be diagnosed as an instance of the withering effect on one's imagination that results from being deeply enamored with a certain picture of the world.

So whether there are any emergent properties of matter is an empirical question to be decided ultimately on the basis of our success in identifying macro-level properties of complex systems with relational complexes of micro-level properties. Now the agency theorist, as we have seen,<sup>9</sup> is committed (on the assumption of a substance monism) to the emergence of a very different sort of property altogether. Instead of producing certain effects in the appropriate circumstances itself, of necessity, such a property enables the particular which possesses it (within a certain range of circumstances) to freely and directly bring about (or not bring about) any of a range of effects. (The number of alternatives genuinely open to an agent will doubtless vary from case to case.) This further commitment leaves the theory's proponent open to a special sort of objection, not applicable to emergentist claims generally: given the unique nature of the sort of property the theory postulates, it is unclear whether it is really conceivable that such a property could emerge from other natural properties. It will be claimed that only a very different sort of substance from material substances, such as is posited by Cartesian dualism,

could possess such a property. (It is noteworthy that many philosophers who discuss the agency theory seem to simply assume that its adherents are dualists.<sup>10</sup>) But given that there is nothing inconsistent about the emergence of an "ordinary" causal property, having the potential for exercising an irreducible causal influence on the environments in which it is instantiated, it is hard to see just why there could not be a sort of emergent property whose novelty consists in its capacity to enable its possessor directly to effect changes at will (within a narrowly limited range, and in appropriate circumstances). And if such a possibility claim is difficult to evaluate on a purely abstract level, it is perhaps more plausible when considered in relation to entities such as ourselves, conscious, intelligent agents, capable of representing diverse, sophisticated plans of action for possible implementation and having appetitive attitudes that are efficacious in bringing about a desired alternative.

The likely reply to this, of course, is that the incoherency of such a view cannot be demonstrated only because we have been given so very "thin" a model to go on. Here, too, I believe, it must be admitted that there is some truth to this charge. Taking the agency theory seriously within a basically materialist framework brings forth a whole host of theoretical problems and issues such as the following<sup>11</sup>: When does a physical system qualify as an "agent"? What structural transformations in the human nervous system would result in long-standing (or permanent) loss of the agent-causal capacity generally? Precisely to what extent is an ordinary human's behavior directly regulated by the agent himself, and to what extent is it controlled by micro-deterministic processes? (Put more generally, how do event- and agent-causal processes interact?) These, however, are obviously empirical matters, requiring extensive advancements within neurobiological science (and advancements favorable, of course, to the agency theorist's commitment to a significant measure of indeterminacy in human behavior). The answers to such questions will not be shown by philosophical work in action theory.

### **III Some contemporary objections to the agency theory**

However, we have yet to examine a few other challenges to the tenability of the agency theory that have been raised in the literature, challenges that clearly are within the province of the philosophical theorist.

Donald Davidson has famously contended that the agency theorist faces an inescapable dilemma, once the question is posed, "how well does the idea of agent causality account for the relation between an agent and his action?"<sup>12</sup> The dilemma that Davidson sees may be expressed thus: either the causing by an agent of a primitive action<sup>13</sup> is a further event, distinct from the primitive action, or it is not.

Suppose first that the agent-causing is a further event. If so, then it is either an action or it is not. If it is an action, then the action we began with was not, contrary to the assumption, primitive. If it is not an action, then we have the absurdity of a causing which is not a doing. Therefore, it seems that we should not say that an agent's causing a primitive action is an event distinct from the action.

Suppose, then, that we grasp the second horn of the original dilemma and maintain that the agent's causing his action does not consist of some further event distinct from his primitive action. Davidson replies:

. . .then what more have we said when we say the agent caused the action than when we say he was the agent of the action? The concept of cause seems to play no role. . . What distinguishes agent causation from ordinary causation is that no expansion into a tale of two events is possible, and no law lurks. By the same token, nothing is explained. There seems no good reason, therefore, for using such expressions as 'cause', 'bring about', 'make the case' to illuminate the relation between an agent and his act. (p.52-3)

Now there are a couple of highly dubious assumptions being made in this passage, but the first response to be made to the putative dilemma is to deny Davidson's assumption that the agency theory maintains that there is an irreducible causal relation between the agent and his

(free) action. For from this perspective, what is most intimately my activity is the causal initiation of my behavior, the causal production of determinate (immediately executive) intentions or volitions. Thus, Bishop writes that on the agency theory,

[t]he action is the existent relation, and may not be collapsed into one of its terms. The object of the agent-causal relation, then, is not the action itself but certain events or sequences of events which, in virtue of their standing in this relation, count as intrinsic to the agent's intentional action.<sup>14</sup>

In the case of an observable bodily movement such as waving my hand, my action consists of the causal relation I bear to the coming-to-be of the state of determinate intention to wave my hand, plus the sequence of events which flow from that decision.<sup>15</sup> How shall we think of the primitive mental action at the core of this larger action? Does it simply consist, as Bishop suggests, of an existent (agent-causal) relation alone? I think that this suggestion is ill-conceived, for the reason that the production of the internal event is not to be identified with the instantiated relation alone, somehow isolatable from its relata, but rather it is the complex event or state of affairs, S's production of e.

Now this, of course, is somewhat at odds with the now conventional analysis of actions as consisting of the events or sequences of events which are produced by an appropriate causal factor. There is good reason to think the conventional analysis is mistaken, however. Consider first the orthodox account of the production of action, viz., the causal theory. On this account, actions are causally produced (at least in part) by desires and beliefs. Such theorists generally claim that there is a sense in which an action may be thought of as produced by its agent on the causal theory - I am the source of my decision to wave my hand in virtue of the fact that my desire to raise my hand (together with certain beliefs) is causally efficacious in bringing that decision about. Thinking of the matter in this way, the event which is my decision, then, is (at least partially) constitutive of an action of mine not solely in virtue of its intrinsic features, but

also in virtue of the fact that it is causally related to me in a certain way. But this is problematic. Is not the production of internal mental events and/or bodily movements an essential part of my activity? If so, then we cannot avoid the conclusion that my primitive action (on the causal account) is to be identified with DB's causing e, where 'DB' is the causally efficacious desire-belief complex.<sup>16</sup>

It will be objected by many that it is simply a mistake to think of the relevant beliefs and desires as components of the action. But whatever unnaturalness this claim appears to possess is to be attributed to the failure of the causal theory to reflect the commonsense view of the etiology of ordinary behavior. If we are inclined to adopt this picture of the springs of action, then since I am active only in virtue of the productivity of properties that constitute my mental state, my being in that state is inseparable from my core activity - that of producing, e.g., a bodily movement.

I will hazard the suggestion that the fact that most action theorists do not individuate actions in this way is in part a result (in some cases indirect) of the influence of Hume's views on causation. Hume and his followers conceive a sequence of events over time as composed of discrete and essentially unconnected elements - "time slices". We may, as a wholly contingent matter of fact, discern various patterns of regularity in the sequences we observe over time, but there are no existent causal relations in nature between events. But if we repudiate this reductionist picture of causality, and allow that causes truly produce their effects, then, as I've just argued, the causal theorist ought to allow that actions are partly constituted by the causal relations that (he maintains) exist between an agent's reasons and resulting behavior.

To return, though, to the task of responding to the dilemma that Davidson attempts to construct, we thus begin by noting that on the agency theory, rather than there being a causal relation between agent and action, the relational complex constitutes the action. Suppose, however, that Davidson were to reformulate his dilemma in terms of the relation between the agent and the event constituents of a primitive (or core) action.<sup>17</sup> The first horn of the dilemma (which assumes that the agent's causing some event is distinct from his action) will then clearly

be idle. But what of the second horn? If we say that the agent's causal activity is identical to his action, is it true, as Davidson asserts, that the concept of cause plays no role in what we assert? That nothing is explained, since we are not connecting the event-constituents of the action to a law?

As far as I can see, Davidson offers absolutely no reason to think we should say this. And, prima facie, such an assertion does seem at least partly explanatory: for if one points to that which causally produced an event, how could one have nonetheless failed to so much as contribute to an explanation of its occurrence? To be sure, such an explanation is far from complete. We have yet to indicate, for example, with what reasons the agent acted as he did. And we have said nothing in specific terms of the sort of nature possessed by the agent, in virtue of which he was capable of bringing about such effects. But one has surely been given something by way of explanation. It seems that Davidson's understanding of the matter here has been clouded by the deleterious effects of Hume, to the effect that explanation of events can only come about through subsumption under a law. But this dogma is essentially tied to the Humean framework, and I take it that there are good reasons for rejecting this.<sup>18</sup> We may safely conclude, therefore, that Davidson's supposed dilemma poses no serious threat to the agency theory.

Another well-known attack on the coherence of the agency theory is made by C.D. Broad (1952).<sup>19</sup> Broad writes:

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? (p.215)

It is far from clear to me just what the difficulty is that Broad takes himself to be pointing out here. It is true that persisting objects such as human agents are not, in the ordinary sense, 'datable' entities, although we may specify the temporal interval through which they exist. But we may, of course, quite unproblematically speak of certain facts being true of an agent at one time that do not hold of him at another. And such is the claim of the agency theorist. Consider, for example, my deliberation a while ago concerning whether to continue working on this paper for another hour or to stop and do something else. After a brief moment of consideration, I formed the intention (at time  $t$ , say) to continue working. According to the agency theory, we may suppose that at  $t$  I possessed the power to choose to continue working or to choose to stop, where this is understood as the capacity to cause either of these mental occurrences. And, in fact, that capacity was exercised at  $t$  in a particular way (in choosing to continue working), allowing us to say truthfully that Tim at time  $t$  causally determined his own choice to continue working. But we needn't, in order to make sense of this, analyze it as the claim (of dubious intelligibility) that a 'datable entity', Tim-at- $t$ , was the occurrent cause of the decision to continue working.

But, you might say, given the fact that your producing your decision occurs at a specific time (and how could it be otherwise?), doesn't it seem appropriate to identify the particular mental state you were in at that time as what was ultimately responsible for that decision (though perhaps in a causally indeterministic fashion, if you like)? What is it about the nature of the causal process as you envision it that prevents us from properly saying this?

My answer is that the alternative wrongly implies that it is whatever is distinctive about the state that the agent was in at the time of his action - distinguishing it from his state just prior to that moment, say - which triggers the action. But while there are various necessary conditions on an agent's producing a decision to  $X$ , these conditions may obtain over a protracted period of time, and so cannot be thought to be themselves causally efficacious (with respect to the decision).

Perhaps underlying Broad's remarks, though, is the thought (made explicit in Ginet 1990) that the proposition that I caused the decision at  $t$  cannot explain why I decided when I did, nor can it explain why I decided as I did. Now this is certainly true, and, we may add, it is further true that analogous questions are answered when we give an event-causal explanation of an event (setting aside complications raised by indeterministic event-causal processes). For causal properties are (ordinarily) such as to immediately give rise to their characteristic effects in the right circumstances, and the effects to which they give rise are characteristic, i.e., it is impossible (either physically or metaphysically) that any other effect should come about in just those circumstances.

But, as I have been at pains to emphasize, agent-causes operate in a different fashion, and corresponding to this is a difference in the way they are involved in the explanation of the effects they produce.<sup>20</sup> An agent-cause does not produce a certain effect in virtue of its very nature, as does an event cause, but does so at will in the light of considerations accessible to him at that time. And so a full explanation of why an agent-caused event occurred will include, among other things, an account of the reasons upon which the agent acted. (The nature of such reasons-explanations, and their degree of explanatory power vis-a-vis fully event-causal explanations, will be considered in the following two sections.)

Yet another instance of an objection which attempts to insist that the agency theory meet standard requirements within an event-causal paradigm which upon reflection are seen to be simply inappropriate in the context of agent-causality is noted by the agency theorist Chisholm:

Our account presupposes that there are certain events which men, or agents, cause to happen. Suppose, then, that on a certain occasion a man does cause a certain event  $e$  to happen. What, now, of that event - the event which is his thus causing  $e$  to happen? We have assumed that there is no sufficient condition for his causing  $e$  to happen. Shall we say it was not caused by anything? If we say this, then we cannot hold him responsible for his causing  $e$  to happen.<sup>21</sup>

I believe that the proper line of response here begins with the observation that the very idea of there being sufficient causal conditions for an agent-causal event is unintelligible. One agency theorist who has endorsed the opposing view is Richard Taylor:

there is nothing in the concept of agency [where this involves an irreducible causal relation between agent and act], as such, to entail that any events must be causally undetermined, and in that sense "free", in order for some of them to be the acts of agents. Indeed, it might well be that everything that ever happens, happens under conditions which are such that nothing else could happen, and hence that in the case of every act that any agent ever performs there are conditions that are causally sufficient for his doing just what he does. This is the claim of determinism, but it does not by itself require us to deny that there are agents who sometimes initiate their own acts. What is entailed by this concept of agency, according to which men are the initiators of their own acts, is that for anything to count as an act there must be an essential reference to an agent as the cause of that act, whether he is, in the usual sense, caused to perform it or not.<sup>22</sup>

We may say that I am free and responsible for some behavior of mine, then, just in case I originate or cause it and am not determined to do so. This would be allowed by Reid and other agency theorists who followed him. Taylor departs from the standard view of agency theorists only in suggesting that the first of these conditions may obtain in the absence of the second. He suggests as a simple, likely case of this sort my grasping my seat tightly while on a ski lift (where my timidity and fright are causally sufficient in the circumstances for my doing so). He notes that it would be odd to say that this is not something I did (compare my concurrent perspiration), and concludes from such examples that it is perfectly intelligible that I should be determined to (agent) cause my own actions. (In Reid's terminology, one may be unfree in the exercise of one's active power.)

Now it is one thing to argue in this way: it is perfectly intelligible that one should be determined on occasion to act as one does; on this theory, one is always the agent cause of one's acts; hence, this theory is constrained to allow for the possibility that an agent is determined to cause his own action. But it is quite another directly to defend the idea of causally determined agent causation against the charge of incoherence. Just how are we to understand the notion of there being a sufficient causal condition for an exercise of active power?

Unfortunately, Taylor himself never tries to spell this out, and he is apparently unaware of the difficulty one faces in trying to do so coherently. Note that what we are to envisage is not that there are sufficient causal conditions for event *e* independently of my causing it, but rather conditions sufficient precisely for the event which is my causing *e* (and only thereby for *e*). (This event is constituted by the holding of a causal relation between myself and the subevent *e*). It is this sort of event for which, Taylor claims, there may be sufficient causal conditions.

For the purpose of evaluating this claim, it will be useful to consider first the case of event causation. The cause of A's causation of B is none other than the cause of A itself.<sup>23</sup> What, then, of S's causation of *e*? There appears to be no way of getting a grip on the notion of an event of this sort's having a sufficient, efficient cause. Because of its peculiar causal structure, there is no event at its front end, so to speak, but only an enduring agent. And there cannot be an immediate, efficient cause of a causal relation (i.e., independently of the causation of its front end relatum). In general, that which is causally produced in the first instance is always an event having a causally simple structure: an object *O*'s exemplifying intrinsic properties *p*<sub>1</sub>, *p*<sub>2</sub>, . . . at time *t*<sub>0</sub>. Causally complex events can also be caused, of course, but only in a derivative way: where they have the form event X's causing event Y, whatever causes event X is a cause thereby of X's causing Y. In the special case of an agent's causing an event internal to his action, however, there is no causally simple component event forming its initial segment, such that one might cause the complex event (S's causing *e*) in virtue of causing it. Therefore, it is problematic to suppose that there could be sufficient causal conditions for an agent-causal event.

If I am right in claiming that it is strictly impossible for there to be sufficient causal conditions for an agent-causal event, we may readily dispose of the objection introduced by Chisholm that if my causing e itself has no cause, then I cannot be responsible for it. For it would appear from the above that no answer could be given to the question of what was the cause of a given agent-causal event, and hence that the question is ill-framed, resulting from a failure to understand the peculiar nature of such an event. In this type of complex event, there is no first sub-event bearing a causal relation to a second. So it seems that the libertarian may acknowledge without embarrassment that events of this type are uncaused.<sup>24</sup>

To support the point I am trying to make here, I want to emphasize the contrast between the scenarios envisaged by the agency theorist and those envisaged by the simple indeterminist. The simple indeterminist claims that a (causally) simple mental event of the proper sort (e.g., a volition), if causally undetermined, is intrinsically such as to be under the control of the agent who is its subject.<sup>25</sup> I have tried elsewhere to give reasons for supposing that the claim is in fact false.<sup>26</sup> Agent-control - the type of immediate control we take ourselves to have over our own actions - is clearly causal in nature.

But now consider an instance of S's causing e. This event is intrinsically a doing, owing to its internal causal structure (i.e., an agent's bearing a direct causal relation to another event). Its very nature precludes the possibility of there being a sufficient causal condition for it (as I argued earlier), being an event which is the agent's causing the event internal to it (e). Now the event e is itself clearly under the control of the agent, since he caused it (directly). But would it not, then, be perfectly absurd to raise a doubt concerning whether the agent controlled his causing e? Indeed, it seems to me that the question of whether the agent has control over this event is ill-framed - it is simply an instance of an agent's exercising direct control over another event.

Chisholm, by contrast, would have the agency theorist maintain that the agent himself causes his agent-causing. It seems that the following line of thought underlies Chisholm's commitment to this perplexing suggestion:

- (1) An agent S bears responsibility for an event x only if S has causally contributed to the occurrence of x.
  - (2) Any instance of an agent's causing an event is itself an event.
  - (3) Agents are responsible for their agent-causings.
- ∴ (4) Agents cause the events which are their agent-causings.

And, of course, if the agent is responsible for an instance of agent-causing by causing it, then we must say that he is responsible for this further event of his causing his agent-causing. And thus in this way we are led (with Chisholm) to fabricate an infinity of simultaneous events. But while (1) seems quite evident when we focus on events that either lack internal causal structure or are constituted by two or more such simple events causally connected to one another, if one allows for the possibility of events that simply are the direct causal activity of agents, then one ought not to suppose that (1) holds with unrestricted generality.

#### **IV Two Objections to Indeterministic Reasons Explanation**

I have yet to address an issue that critically bears on the viability of the agency theorist's general project of providing an adequate theoretical framework for understanding how free agency operates. We explain the actions of ourselves and others around us by citing or ascribing reasons for which the action was performed. How do reasons figure into the performance of actions as the agency theorist conceives them? It is astonishing to me to see how often critics of the agency theory make the mistaken assumption that the agency theory is either incompatible with reasons-based accounts of action or is advanced as an independent alternative to such accounts. This leads naturally enough to the conclusion that it is simply confused<sup>27</sup> or explanatorily superfluous<sup>28</sup>. In this section and in the one that follows, I will try to show how agent causality plays a necessary role in reasons explanations, once we abandon the causal theory's model of reasons as influencing actions by causally producing them.

I will begin by considering two recent objections of a highly general character against the possibility of a satisfactory account of non-causal reasons explanations. The first of these is Galen Strawson's claim that the indeterminist's conception of of an agent as acting in view of prior motives while not being determined by them ineluctably leads to a vicious regress. For, he claims, we can conceive of an agent sitting in detached judgment on the matter of whether to act in accordance with motive X or motive Y only if he has some further desires or principles of choice that decisively inclines him in one of these directions. But if this is the case, then the agent is self-determining in making his choice only if he is somehow responsible for the presence of those further factors, which requires his having chosen to be that way. . .<sup>29</sup>.

Strawson is not alone in holding that the libertarian is unwittingly committed to this (problematic) picture. The same suggestion was colorfully made by Leibniz:

One will have it that the will alone is active and supreme, and one is wont to imagine it to be like a queen seated on her throne, whose minister of state is the understanding, while the passions are her courtiers or favourite ladies, who by their influence often prevail over the counsel of her ministers. One will have it that the understanding speaks only at this queen's order; that she can vacillate between the arguments of the ministers and the suggestions of the favourites, even rejecting both, making them keep silence or speak, and giving them audience as it seems good to her. But it is a personification or mythology somewhat ill-conceived.<sup>30</sup>

But while we may wholeheartedly agree with Leibniz's assessment of this conception as "somewhat ill-conceived", we should also reject the suggestion that the libertarian must be assuming (if the account is to avoid positing fortuitous, irrational choices) that the agent has further, second-order reasons that explain why he chose to act in accordance with one set of motives rather than another. Consider a scenario in which an agent is deliberating between two courses of action X and Y, each of which has considerations in its favor. (I will refer to these

sets of considerations as {X} and {Y}, respectively.) Suppose further that the agency theory is correct and the agent herself brings about the decision to take option X. The question, "Why did the agent perform that action?", is meaningfully answered by citing {X}, even though these reasons did not produce the agent's decision, and she could have chosen differently in those very same circumstances. In citing {X}, we are explaining the motivating factors that were in view when the agent made a self-determining choice. It is not necessary to try to ascend to a level of second-order reasons (for acting on first-order reasons) in a desperate bid to render this conception of action intelligible.

Perhaps underlying Strawson's charge is the belief that an action would be irrational or at least arbitrary (in a pejorative sense of that term) if, at the time of acting, the agent did not believe that her reasons decisively favored the course of action chosen, that she had reasons for performing X rather than Y. The first thing to notice here is that even if we were to accept this, it does not clearly imply that the agent-causationist model must be mistaken. For we should then say, in any given case, that while the agent has it in her power to choose any of a range of alternatives, only one choice would be rational from the standpoint of her own reasons. Why would it still be "rationally-speaking random", as Strawson puts it (in a portion of the text not quoted here), if the agent makes the preferable choice? And, furthermore, don't we sometimes make irrational decisions? It is open to Strawson to accept my claim that choices of the most preferable option would be rational, but then suggest that the power the agency theory confers on free agents is worthless. For it is nothing but the power to make irrational decisions, and who wants that? I do not accept the suggestion that there is no value in an agent's freely choosing to be (for the most part) rational. But we can say something further. And that is that many situations of choice simply do not point to one course of action as "the thing to do" in the circumstances, as being preferable to all the rest.<sup>31</sup> Moral choices are commonly of this sort, but it is not limited to these. And if this claim is right, then there will be situations in which the agency theory confers a power on agents beyond that of determining whether they shall act rationally or irrationally.

In responding to Strawson's contention that the indeterministic aspect of the agency theory leads to a regress of reasons, I have begun to stray into the territory of the second, related objection to indeterministic reasons-explanation that I want to consider, and so it is appropriate now to make this objection explicit. Suggested by various remarks in Kane (1989),<sup>32</sup> the objection I have in mind may be put thus: Any genuinely explanatory response to the question, "Why did S do X?", will ipso facto be an answer to the question, "Why did S do X rather than any of the available alternatives?", and a proper answer to the latter of these must incorporate all the relevant psychological features of the agent at the moment of choice. This requires some elaboration. The question, "Why did S do X rather than, say, Y or Z?", might be interpreted as simply a request for the reason that motivated S's making the choice S made (as opposed to reasons there may have been for any of the alternatives to X). When the question is construed in this way, however, we needn't cite all the considerations before S's mind at the time of the decision, but only those that provided a motive for doing X. But the sort of reading of this question that Kane has in mind is a much stronger one, which requests an account of why it was necessary that S do X in those circumstances rather than any of the alternatives. And citing whatever motive(s) there were for doing X at that time is clearly insufficient for this purpose. Rather, we need an account that implies that those motives were enough to tip the balance in favor of the actual outcome, as against its competitors.

Consider the following remarks in Kane (1989):

How can we explain either outcome, should it occur, in terms of exactly the same past?

If we say, for example, that the agent did [X] rather than [Y] here and now because the agent had such and such reasons or motives and engaged in such and such a deliberation before choosing to act, how would we have explained the doing of [Y] rather than [X] given exactly the same reasons or motives and the same prior deliberation? (p.228, emphasis added)

In simply assuming that an explanation of the action will cite all the salient psychological features of the agent at the time of his decision, Kane is clearly presuming that there is only one type of adequate explanation of a choice, the type that explains why only that choice could have been made at that point in the agent's psychological history. But this is unsupported. The agency theorist may cheerfully concede that explanations of that sort are precluded by actions that are described by his theory - i.e., explanations that cite factors which could put an observer in a position to predict outcomes with certainty. And though we may grant that explanations of that sort are highly desirable for scientific purposes (among others), no reason has been given why we cannot allow explanations that account for an occurrence by characterizing it as the freely initiated behavior of an agent motivated by such and such a reason.<sup>33</sup>

The element of causal initiation is critical, I think, to the viability of this alternative explanatory framework. Some philosophers have failed to see that the prior presence of consciously-considered reasons and agent-causal initiation are each necessary components in the agency theorist's explanatory scheme, and so have drawn the conclusion that the role of reasons in explaining actions obviates appeal to agent causation. Thus, Goetz (1988), for example,<sup>34</sup> writes:

. . .if the reasons for which an agent acts help explain her freedom and responsibility with respect to that action, and her causing of her action can only be explained by appeal to the reason for which she acts, it is clear that the agent's causing of her action cannot help explain how it is that the agent is free and responsible with respect to her action. Any explanatory power which the causation by the agent of her action might have would have to be derived from or parasitic upon the explanatory power of the reason she has for performing that action. Thus, not only is it the case that agent-causation cannot help explain an agent's performance of a free action, but also it is not needed for this explanatory role, once the agent's reason for performing that action has been invoked to explain it. (p.310, emphasis added)

The sentence I have highlighted in the above passage involves a mistaken claim. It is doubtful that we can form a conception of an agent's causing an event internal to his action without his having any sort of pro-attitude towards that action, and so to that extent the agent's causing the component event is dependent on the reason he has (or, his having a reason) for acting in that way. Nonetheless, the relative dependency of reasons and agent-causal initiation with respect to explanatory power is precisely the reverse of what Goetz suggests. For the agent's free exercise of his causal capacity provides a necessary link between reason and action, without which the reason could not in any significant way explain the action. It allows us to claim that the reason had an influence on the production of the decision, while not causing it. Were we to remove the element of causal production of decision altogether, and simply claim that the decision was uncaused, then noting the fact that the agent had a reason that motivated acting in that way would not suffice to explain it (as Davidson has famously argued). For in that case any number of actions may have been equally likely to occur, and the agent would not have exercised any sort of control over which of these was actually performed (either via the efficacy of his reasons or in the direct fashion suggested by the agency theory). And it seems sufficiently obvious that where there are no controlling agents/ factors of even a relatively weak, indeterministic sort, there can be no explanation of the occurrence.

It will be observed that the crucial claim I am making here is that any genuine explanation of an occurrence must involve an account of how that occurrence was produced. It has often been thought that, given this requirement, the only way in which reasons can play a role in the explanation of an action is by functioning as the central features of a set of conditions that determine the action. One alternative to this is to suppose that reasons cause actions without determining them. I do not deny that this is a viable indeterministic account of reasons explanation rival to the one I am offering here. But while it provides for the possibility of reasons explanation, I think it must be rejected ultimately on the grounds that it fails to show how it can be up to an agent to determine which among a range of possible courses of action he

will actually undertake.<sup>35</sup> If I am right in supposing this, then the only account of reasons explanation that is consonant, in the final analysis, with a picture of free and responsible agency is the one suggested by the agency theory.

## V An Account of Reasons Explanation

So far, however, I have only spoken impressionistically of the sort of reasons explanation appropriate to the agency theory. I will now attempt to give a more careful account by laying out conditions sufficient for the truth of each of two general sorts of ordinary reasons explanation. (What I say about these cases is readily adaptable to other sorts, such as explanation by a prior intention.)

The first sort that I want to consider involves explaining action by reference to a prior desire that  $\emptyset$ , where this is construed broadly (and beyond everyday usage) as including any kind of "pro-attitude" or positive inclination towards the state of affairs  $\emptyset$ . The following general conditions seem to me to suffice for the truth of an explanation of an action in terms of an antecedent desire:

S V-ed then in order to carry out her antecedent desire that  $\emptyset$  if:

- (i) prior to this V-ing, S had a desire that  $\emptyset$ , and believed that by V-ing, she would satisfy (or contribute to satisfying) that desire, and
- (ii) S's V-ing was initiated (in part) by her own self-determining causal activity,<sup>36</sup> and
- (iii) concurrent with this V-ing, S continued to desire that  $\emptyset$  and intended of this V-ing that it satisfy (or contribute to satisfying) that desire.

Condition (iii) is necessary<sup>37</sup> because were I to cease to have the original desire and act for a completely different reason, it clearly would not have a genuinely explanatory role to play. It also handles cases in which I continue to have the desire but it is not the reason for which I act (and hence I don't intend of my action that it satisfy that desire).

This third condition is an adaptation of the central component of Ginet's account<sup>38</sup> of reasons explanation, although there is an important difference in how it functions in our overall accounts. I am in agreement with Ginet that the part of the explanation that involves a connection between the prior desire and the present intention need not be causal in nature (apart from the causal connections involved in continuing to have the desire), but may, rather, be wholly internal (similarity of content) and referential. If it is my purpose or intention in V-ing that I carry out a prior desire that  $\emptyset$ , then the prior desire may figure in the explanation of this action even if it does not constitute part of a set of conditions that causally produce the action. Contra Ginet, however, this will be the case only if the non-causal connection between desire and intention is coupled with some other, appropriate sort of factor that produces or initiates the action, viz., the agent herself (hence, the necessity of condition (ii)).

In discussing Ginet's simple indeterminist, non-causal account of reasons explanations (which lacks anything analogous to my condition (ii)), Lawrence Davis writes:

. . . "she opened the window in order to let in fresh air" only if she opened the window because she believed she would or might let in fresh air thereby. And this "because" must be causal - nomic - else I do not see a plausible distinction between [this sentence] and

(1') She opened the window knowing she would let fresh air in thereby.

. . . If I am right that something causal is needed, . . . [then] Ginet has not shown that undetermined acts can be explained in terms of their antecedents.<sup>39</sup>

I think that Davis is right in supposing that "something causal is needed" to make possible an explanatory link between antecedent reason and action, but that causal element needn't be a nomic connection between reason and action. The agency theory provides a coherent framework in which reasons can influence the production of an action without themselves forming part of a causally sufficient condition for the action.

I might note that there may be further factors that enter into the explanation of my V-ing. Suppose my prior desire was relatively indeterminate with respect to when it should be realized. There will often be certain considerations or other factors at the time of acting that elicited my action (by suggesting that this was a particularly opportune time to satisfy the desire), and these will certainly figure in a full explanation of my action. But, by the same token, it's not obvious that there needs to be such environmental stimuli. Perhaps I am only concerned that I act within a certain time frame, and any particular moment is as good as any other. In such a case, there may not be an explanation of why I acted just then (rather than at some other time).

The set of sufficient conditions for an explanation of action by prior reasons just given are consistent with the agent not having a clear preference for the action performed over any available alternative. However, we often do act on such preferences; in such cases, we can explain (in terms of antecedent reasons) not only why the agent V-ed, but also why she V-ed rather than doing something else instead. Can we give non-deterministic sufficient conditions for the truth of such explanations, similar to those sketched above? I think that we clearly can. Consider the following:

S V-ed then rather than doing something else because she preferred V-ing to any alternative if:

- (i) prior to this V-ing, S had a desire that  $\emptyset$ , and believed that by V-ing, she would satisfy (or contribute to satisfying) that desire, and
- (ii) S preferred V-ing as a means to satisfying the desire that  $\emptyset$ , and also preferred satisfying  $\emptyset$  over the satisfaction of any other desire, and
- (iii) S's V-ing was initiated (in part) by her own self-determining causal activity, and
- (iv) concurrent with this V-ing, (a) S continued to desire that  $\emptyset$  and intended of this V-ing that it satisfy (or contribute to satisfying) that desire, and (b) S continued to prefer V-ing to any alternative action she believed to be open to her.

It is quite consistent with the antecedent circumstances expressed in these conditions that S have failed to V at that time. She might, for example, have come to prefer on reflection some alternative (or have ceased to desire that  $\emptyset$  altogether), or she might have decided to continue seeking out further relevant considerations, or, finally, she might have simply succumbed to some temptation despite her continuing to believe that V-ing represented the best course of action open to her (thereby exhibiting the phenomenon of "weakness of will").<sup>40</sup>

It might be claimed, however, that the fact that our set of conditions does not rule out these possibilities goes to show that they are not truly sufficient for the truth of explanations of why an agent performed a particular action rather than any alternatives she had considered. (i)-(iv) must be supplemented with conditions that rule out the possibility of these alternative scenarios. Only then, it will be claimed, will we have adequately explained why S performed the action she did, rather than some other action.

But while similar charges have often been made by critics of libertarianism, as best I can see, there are no good reasons to accept them. We may suppose that it is a wholly contingent matter of fact that none of the alternative scenarios I envisaged occurred, that the prior circumstances did not necessitate their non-occurrence. How does this show our set of conditions to be inadequate? If what we are seeking to explain is why a particular action was in fact undertaken rather than some other, as opposed to why the action had to occur, why is it not enough that we refer to those antecedent reasons the agent had for preferring the chosen action over the alternatives, reasons the agent continued to have at the time of the action and which she intended to satisfy in performing it? Providing such an explanation clearly makes it teleologically intelligible that the agent chose to perform that action rather than any of the others, though it does not imply that no other action could have occurred in just those circumstances. Therefore, I cannot see why it should be thought that it fails genuinely to explain the action in any meaningful sense - unless, again, the critic is failing to note the difference between agent-causal and entirely non-causal reasons explanations, a difference that is embodied in my third condition above.

## VI Is Agent Causation Distinguishable From Mere Randomness?

The final objection to the agency theory that I will consider here is epistemological in nature: It seems that it is impossible, in principle, for us ever to know whether any events are produced in the manner that the agency theory postulates. For such an event would be indistinguishable from one which was essentially random, not connected by even probabilistic laws to events preceding it.<sup>41</sup> (Alternatively put, the objection claims that we could never know whether the unique sort of property or properties which give rise to active power is instantiated.)

However, if my earlier contention that simple indeterminism is incompatible with genuine reasons explanations of action is correct, then I believe that the present objection must be judged mistaken. The simple indeterminist supposes that (in many cases) an agent's decision is not the outcome of any determinative causal influence - neither the agent's prior reasons, as on the causal theory, nor simply the agent qua agent (as on the agency theory). I claimed, though, that reasons explanations require a mechanism of control that 'hooks up', so to speak, the agent's reasons and consequent decision (and action). On the causal theory, this is supplied by an event-causal relation between the decision and matching reason(s). On the agency theory, an agent's capacity directly to produce a decision in the light of consciously-held reasons fills the bill. We cannot simply appeal, as, for example, Ginet (1989) does, to internal (and referential) relations between concurrent intention and prior motives, on the one hand, and that same concurrent intention and the decision (or action), on the other. Without the mediation of a (necessarily causal) 'mechanism of control', prior motives cannot explain a decision, even though (as it happens) they may coincide with it.

Returning now to the objection under consideration, let us suppose that our knowledge of natural processes were to progress to such a point as to provide unmistakable evidence of significant indeterminism in the nature of ordinary human action. Would we have no reason, in such an eventuality, to prefer the agent-causal hypothesis to that of simple indeterminism? Surely not. Surely it would be preferable to adopt a theory of action in virtue of which our

reasons-based explanations could remain largely intact. And it seems that, in such a scenario, only the agency theory would allow this. Furthermore, given a detailed knowledge of neurophysiological processes, we could go beyond the bare postulation of the appropriate property (i.e., one on which the power to cause directly any of a certain range of alternative events supervenes). We could explain in some detail, for instance, the systemic conditions under which such a property is instantiated, as well as the subtleties of its interplay with other causal processes involved in the production of behavior.

Thus, the employment of the concept of active power is not irremediably at odds with the attempt to give a scientific account of natural processes, including human behavior (as is sometimes alleged). The use of this concept in explaining human behavior is consonant with scientific methodology and could, in principle, be mapped onto other explanatory theories concerning biological subsystems of the human organism. It does run counter to the general program of micro-reductive explanation, which has been highly successful in other contexts. But this, it surely must be recognized, is simply a research strategy. Given its explanatory potential, it obviously should be pushed as far as it can go in the understanding of human behavior. (And there is a further reason that agent-causal mechanisms should be appealed to in theoretical accounts only after the alternatives have been exhausted: we simply cannot know in advance the details of how event- and agent-causal processes interact, nor the precise sorts of circumstances in which agent-causal processes do not figure at all in the production of behavior.) But if limits of the right sort persist, I see no reason that explanatory theories invoking the concept of agent causality should not be adopted.<sup>42</sup> The alternative - to regard much of our behavior as without explanation (save for the fact that it falls within certain parameters) - is simply not credible.

This reply to the charge that we could never have reasons for preferring the agent-causal form of explanation to that of causal randomness may be bolstered by a simple appeal to how things seem to us when we act. It is not, after all, simply to provide a theoretical underpinning for our belief in moral responsibility that the agency theory is invoked. First and foremost (as I

suggested at the outset), the agency theory is appealing because it captures the way we experience our own activity. It does not seem to me (at least ordinarily) that I am caused to act by the reasons which favor doing so; it seems to be the case, rather, that I produce my decision in view of those reasons, and could have, in an unconditional sense, decided differently. This depiction of the phenomenology of action finds endorsement not only, as might be expected, in agency theorists such as Reid, Campbell, and Taylor, but also in determinists such as Bradley<sup>43</sup>, Nagel<sup>44</sup>, and Searle<sup>45</sup>, and in Ginet's "actish phenomenal quality"<sup>46</sup>. If these largely similar accounts of the experience of action are, as I believe, essentially on target, then it is natural for the agency theorist to maintain that they involve the perception of the agent-causal relation. Just as the non-Humean is apt to maintain that we not only perceive, e.g., the movement of the axe along with the separation of the wood, but the axe splitting the wood (Madden and Harré, 1975, pp.49-51), so I have the (putative, at any rate) perception of my actively and freely deciding to take Seneca St. to my destination and not Buffalo instead.<sup>47</sup> Such experiences could, of course, be wholly illusory, but do we not properly assume, in the absence of strong countervailing reasons, that things are pretty much the way they appear to us? I will not delve into this further epistemological issue here, my concern being that of descriptive metaphysics, but I will note that skepticism about the veridicality of such experiences has numerous isomorphs which, if accepted, appear to lead to a greatly diminished assessment of our knowledge of the world, an assessment that most philosophers resist.<sup>48</sup>

---

<sup>1</sup> As has recently been suggested, for example, by David M. Armstrong (What is a Law of Nature? Cambridge: Cambridge University Press, 1984) and Michael Tooley (Causation: A Realist Approach Oxford: Clarendon Press, 1987).

<sup>2</sup> Causal Powers: A Theory of Natural Necessity (Oxford: Basil Blackwell).

<sup>3</sup> "Causality and Properties", in Identity, Cause and Mind (Cambridge: Cambridge University Press).

<sup>4</sup> As Reid clearly saw, the notion of a particular actively (or agent-causally) bringing about an effect is intelligible only on the supposition that the particular be an agent capable of representing possible courses

---

of action to himself and having certain desires and beliefs concerning those alternatives. (The reader is invited to try to form the conception of an object constituting a counterexample to this claim.) This simple observation is sufficient to dismiss the derisive query of Watson ("Free Action and Free Will", Mind 94, 1987, pp.145-72) as to whether it is conceivable that spiders should turn out to be "agent-causes in Chisholm's sense" (p.168).

<sup>5</sup> Searle (Minds, Brains, and Science. Cambridge, MA: Harvard University Press, 1984), p.94.

<sup>6</sup> Reid would not have recognized indeterministic natural processes as an alternative to causal necessity. Hence, his claim that there is no reason to suppose that intelligent material substances (if such there be) could not be capable of free action is a defense of the possibility of a material system's exercising agent-causality.

<sup>7</sup> Reid, Essays on the Active Powers of the Human Mind (Cambridge: MIT Press, 1969), p.367.

<sup>8</sup> For further details, see my "Emergent Properties" (American Philosophical Quarterly, forthcoming); cf. Brian McLaughlin, "The Rise and Fall of British Emergentism", in Beckermann, A., Flohr, H., and Kim, J., eds., Emergence or Reduction?: Essays on the Prospects of Nonreductive Physicalism (Berlin: Walter de Gruyter, 1992).

<sup>9</sup> Because agent-causality is a distinct species of causation, only an emergent type of property could enable its occurrence.

<sup>10</sup> Two examples among many are Honderich (A Theory of Determinism, Oxford: Oxford University Press, 1988) and Levison ("Chisholm and 'the Metaphysical Problem of Human Freedom'", Philosophia 8, 1978, pp.537-541).

<sup>11</sup> Actually, an adherent of a viable dualist version of the agency theory would have to answer much the same sort of questions as those suggested above.

<sup>12</sup> "Agency", in Essays on Actions and Events (Oxford: Oxford University Press, 1980), p.52.

<sup>13</sup> I.e., an action that one performs without doing anything else in order to perform it. Theories of action differ on the question of the class of actions that fall within one's repertoire of primitive or basic actions, but plausible candidates include decisions and simple bodily movements.

<sup>14</sup> "Agent-causation", Mind 92, 1983, p.71.

<sup>15</sup> It would be a mistake, I think, to characterize a decision of the action-triggering type as simply the occurrence of an event which is, as I've been putting it, the coming-to-be of a state of intention to  $\emptyset$ . While this construal is natural, of course, on causal theories of action, the

---

agency theory conceives of the activity of decision-formation as centrally involving the agent causation of such an event. Consequently, the formation of decision is most properly defined as a complex state of affairs consisting of the agent's bearing a causal relation to a causally-simple mental event (which, I have suggested, we may take to be the coming-to-be of a state of intention to  $\emptyset$ ). Some agency theorists have spoken of "causing one's own decision"; I suggest that they are best interpreted as expressing the above idea in shorthand. In what follows, I will make use of this convenience also from time to time, and the reader is to interpret such statements in the preceding manner.

<sup>16</sup> Fred Dretske has argued for just this claim in Explaining Behavior, Ch.2.

<sup>17</sup> Bishop (1983), pp.72-3, suggests this reformulation of Davidson's argument. However, I have differed with Bishop's interpretation of Davidson's remarks in posing the second horn of the dilemma, and consequently my response to it takes on a different form from his.

<sup>18</sup> Most fundamentally, I fail to see how merely indicating that an event falls under a pattern of regularity - no matter how "lawlike" the formal characteristics of that pattern may be - is, in and of itself, explanatory. It is only by indicating something concerning the causal mechanism(s) at work (as a broadly realist position understands this notion) that genuine explanation can be accomplished. Most Humeans, of course, do not see the matter this way.

<sup>19</sup> "Determinism, Indeterminism, and Libertarianism", in Ethics and the History of Philosophy (London: RKP). Broad's argument has been endorsed by Ginet (On Action, Cambridge: Cambridge University Press, 1990, pp.13-14), although Prof. Ginet has told me in conversation that he no longer feels certain that the apparent difficulty Broad raises is decisive. And a similar (though, to my mind, less clear) sort of objection to Reid's agency theory is raised by Baruch Brody in his introduction to an edition (1969) of Reid's Essays on the Active Powers.

<sup>20</sup> As I noted above in responding to Davidson.

<sup>21</sup> "Reflections on Human Agency", Idealistic Studies 1, 1971, p.40.

<sup>22</sup> Action and Purpose (Englewood Cliffs: Prentice Hall, 1966), pp.114-15.

<sup>23</sup> Assuming, that is, that what we are after is the 'triggering' cause of the event, rather than what Fred Dretske calls a 'structuring' cause - roughly, that which establishes a causal pathway between two objects or systems so that when the first is operated upon (by the triggering cause) in the right manner, it brings about a result in the latter.

---

<sup>24</sup> Of course, there will be a large number of necessary causal conditions for the occurrence of any instance of an agent's directly causing some internal mental event at a particular time *t*. (And, hence, where any of these are absent, a sufficient condition for the nonoccurrence of such events.) Many of these will have to do with the internal state of the agent prior to *t*. To note only the most obvious such conditions, for an agent to cause, say, his decision to immediately engage in Ø-ing, the option must be one that is accessible to his conscious awareness, he must believe it to be within his power, and, it would seem, he must have some positive inclination to Ø. There will of course also be numerous conditions in terms of the structural constitution of the agent's neurophysiological system. It is evident from our acquaintance with pathological cases that very subtle forms of malfunctioning can vitiate or even negate altogether the agent's capacity to act with a normal degree of autonomy.

<sup>25</sup> See, e.g., Carl Ginet, "Reasons Explanation of Action: An Incompatibilist Account", Philosophical Perspectives 3, 1989, pp.17-46.

<sup>26</sup> "Indeterminism and Free Agency: Three Recent Views", Philosophy and Phenomenological Research 53(3), 1993, pp. 499-526.

<sup>27</sup> See Honderich (1988), pp.196-97.

<sup>28</sup> See Stewart Goetz, "A Noncausal Theory of Agency", Philosophy and Phenomenological Research 49, 1988, pp.303-16.

<sup>29</sup> Freedom and Belief (Oxford: Oxford University Press, 1986), pp.53-4.

<sup>30</sup> Theodicy (LaSalle: Open Court Publishing Co., 1985), p.421.

<sup>31</sup> Helpful discussions of such choice scenarios are found in Robert Kane's Free Will and Values (Buffalo: SUNY Press, 1985) and Peter van Inwagen's "When Is the Will Free?", reprinted in this volume.

<sup>32</sup> "Two Kinds of Incompatibilism", Philosophy and Phenomenological Research 50, 219-254. See, e.g., pp.227-8. Kane's remarks in this connection are endorsed by Richard Double in The Non-Reality of Free Will (New York: Oxford University Press).

<sup>33</sup> In a recent discussion, Randolph Clarke helpfully calls attention to the fact that a strong case has been made (quite apart from the special case of reasons explanation) by contemporary philosophers of science that explanation of an event needn't involve showing why it rather than any other possible outcome obtained. See his "A Principle of Rational Explanation?" (Philosophy and Phenomenological Research 30(3), 1992, pp. 1-12) and the articles he cites there.

<sup>34</sup> A similar claim is made by Irving Thalberg in "How Does Agent Causation Work?" (pp. 213-38 of Brand and Walton, eds. Action Theory, 1976; see p. 234f.).

---

<sup>35</sup> I argue this claim at length in O'Connor (1993). See also Ch.4 of P. van Inwagen's An Essay on Free Will (New York: Oxford University Press, 1983).

<sup>36</sup> As I suggested above, I am inclined to term the agent-causal event (S's causation of e) a "decision", the event component of which is the-coming-to-be-of-an-action-triggering-intention-to-V-here-and-now. It is plausible to take it that the intention one has concurrent with the full performance of the action (required in condition (iii) in the text) is a direct causal consequence of the action-triggering-intention that is directly brought about by the agent.

<sup>37</sup> More precisely, some condition or other that is more or less like condition (iii) is needed to give a sufficient condition for acting in order to carry out a prior intention.

<sup>38</sup> Ginet (1989).

<sup>39</sup> Review of Ginet's On Action (Mind 100 (3), 1991, p.393.)

<sup>40</sup> Cf. Ginet (1990), p.149.

<sup>41</sup> An objection along these lines is presented by Alvin Goldman in A Theory of Human Action (Englewood Cliffs, N.J.: Prentice Hall, 1970).

<sup>42</sup> For discussion of the possible use of the concept in the social sciences, see p. 84 of John Greenwood, "Agency, Causality, and Meaning", Journal for the Theory of Social Behavior 18(1), 1988, pp.95-115.

<sup>43</sup> "Free Will: Problem or Pseudo-Problem?", Australasian Journal of Philosophy 36, 1958, pp.33-45.

<sup>44</sup> The View From Nowhere (Oxford: Oxford University Press, 1986), Ch.7.

<sup>45</sup> 1984, op. cit.

<sup>46</sup> Ginet, 1990.

<sup>47</sup> Donagan, surprisingly, is an agency theorist who professes to find the notion of directly perceiving one's causal activity unintelligible (Choice: The Essential Element in Human Action. London: RKP, 1987, pp.181-2). Judging by his remarks there, however, I suspect that he would reach a similar verdict with respect to the notion of perceiving certain instances of event-causal activity.

<sup>48</sup> Many people have given me helpful suggestions and criticisms of material presented in this paper. I wish to acknowledge in particular the help of Randolph Clarke, Mark Crimmins, Norman Kretzmann, Al Plantinga, Dave Robb, Sydney Shoemaker, and, especially, Carl Ginet.