

## **Dualist And Agent-Causal Theories**

Timothy O'Connor

in Robert Kane, ed., *Oxford Handbook on Free Will*  
(Oxford University Press, 2001), 337-355

### I Introduction

This essay will canvass recent philosophical accounts of human agency that deploy a notion of 'self' (or 'agent') causation. Some of these accounts try to explicate this notion, whereas others only hint at its nature by way of contrast with the causality exhibited by impersonal physical systems. In these latter theories, the authors' main argumentative burden is that the apparent fundamental differences between personal and impersonal causal activity strongly suggest mind-body dualism. I begin by noting two distinct, yet not commonly distinguished, philosophical motivations for pursuing an agent-causal account of human agency. In the course of discussing the accounts that some philosophers have developed in response to these considerations, I reconsider both the linkage of agent causation with mind-body dualism and its sharp cleavage from impersonal (or 'event') causation.

### II Mechanism, Teleology, and Agency

A central motivation for many philosophers who espouse an agent-causation-based account of human action is the thesis that mechanistic explanations of a sort found in the physical sciences and purposive explanations of the sort typically applied to human action are in general mutually exclusive: it could not be true, say, that a neurophysiological account (referring only to electrochemical and biological properties) and a purposive account (referring to an agent's beliefs, desires, and intentions) are complementary, true accounts of the very same phenomena - say, the agent's picking up a book.

One indirect line of argument for this thesis is suggested by the 'explanatory exclusion problem' developed over numerous articles by Jaegwon Kim. (For a recent statement, see Kim 1998.) Kim directs his argument against the nonreductionist-physicalist position that mental events and processes are distinct from, yet wholly 'realized by,' physical processes.

He tries to show that such a view cannot coherently account for the causal efficacy of the mental, assuming (as is plausible on the view) that there are no causal factors beyond the physical (the 'causal closure of the physical') and that mental causes do not systematically overdetermine events caused by physical factors. Kim's argument, if successful, forces one to either assume an outright identity of sorts between mental events and specific physical events - something many philosophers regard as highly implausible - or to move toward a more robust sort of dualism, entailing the falsity of the closure of the physical. While taking the latter option concerning the ontology of mental states does not tell in favor of any specific account of agency, it does open up the possibility of supposing that the explanation of actions in terms of mental factors is quite different in kind from the general mechanistic form of explanation universally applied in the purely physical domain.

An argument directly for the distinction in form between these two sorts of explanation that was popular in the 1960s was articulated by Norman Malcolm (1968): mechanistic forms of explanation posit merely contingent lawful connections between events, whereas purposive explanations posit necessary connections between desires or intentions and actions. Therefore, instances of the latter cannot be explained by instances of the former, conceived as more fundamental in both ontological and explanatory terms. It is concluded that since purposive explanations are not illusory, mechanistic explanations are insufficient for the understanding of human agency.

A second direct line of argument that purposive explanations differ in kind from mechanistic ones is of special interest here, as it aims to show the necessity of a specifically agent-causal account of purposive action. In brief, the argument is this: the only plausible way to make out purposive explanations as special cases of mechanistic ones is to hold that desires, intentions, and beliefs are important and salient antecedent causes of action. However, this 'causal theory of action,' championed by Donald Davidson in his influential (1963), faces the problem of wayward causal chains: possible scenarios which satisfy the causal theory's requirement, but which seem not to be cases of purposive actions, due to the

wayward or deviant way in which the agent's motivational states give rise to the action. For example, suppose David desires to kill Ser-Min by poisoning his tea. His desire to do so makes him very nervous, so much so that it causes him to spill the poison into the tea. Here, David's desire causes an action of the intended sort, but he did not act intentionally, or with the purpose of poisoning the tea. Such examples show that the causal theorist must refine her account, specifying the right type of way that motivating factors cause actions that are genuinely intentional. This has proven a difficult task, prompting Roderick Chisholm (1966), John Bishop (1983, 1986), and especially Richard Taylor (1966) to contend that any adequate account of intentional action must include the notion of the agent's causing the action as a primitive notion. Not many have been convinced by this pessimism, however, as a small cottage industry has sprung up to remedy the defects in the simple causal account. To date, the most sophisticated attempts by causal theorists are Bishop himself (1989) and Alfred Mele (1992).

### III Mechanism and Freedom of Action

People act intentionally throughout their waking lives. Whether or not they do so freely is a further substantive matter, one that depends on the nature of their control over their own actions. The concept of action is distinct from that of free action, and it is not obvious that a good way to understand freely performed action is to develop a set of plausibly sufficient conditions for action and then to add a further freedom condition. For perhaps there are more than one interestingly different ways that the concept of intentional action might be satisfied, but some of these do not admit of freedom variants. To speak more concretely: many theorists of action have found it plausible (pace Taylor and the early Bishop) that a suitably nuanced causal theory can provide a set of sufficient conditions for intentional action. (Some have also thought that their favored accounts provide necessary conditions as well, but that is rarely argued with any care.) Suppose that they are right. It may be, however, that a picture of actions as causal products of appropriate motivational states of the agent

cannot provide an account of the sort of freely chosen activity we typically ascribe to ourselves, a sort that grounds moral responsibility and the significance we accord to some of our achievements.

Some agent causation theorists are best interpreted in this manner. Agent causation is a necessary feature of freely chosen activity, even though there may be possible forms of intentional activity that lack it altogether. (C.A. Campbell 1967, John Thorp 1980, Alan Donagan 1987, Randolph Clarke 1993, 1996, and O'Connor 1993, 1995, 2000 explicitly take this view, while Chisholm vacillates in his early essays. Taylor, as already noted, propounded agent causation as a feature of all intentional action, as does Godfrey Vesey 1987, William Rowe 1991, and Richard Swinburne 1986, assuming Swinburne is in fact an agent causationist - on which question see note 4 below.) Indeed, it may be that while some of our own actions are agent-causal in character, others are not. For the remainder of this essay, I will focus solely on this freedom-based motivation for developing an agent-causal account of agency.

Agent causationists have generally also been incompatibilists, holding that freedom of action and causal indeterminism are incompatible. Now a causal theorist of action can likewise endorse incompatibilism. While an intentional action, on her account, involves one's reasons causing one's actions, the causal connection need not be deterministic. So she might suppose that a freely performed action is one in which the agent's reasons are salient parts of an indeterministic causal condition generating the action. The condition causes the action, but it need not have done so. It might have been, in those very circumstances, that one of the agent's competing desires (and corresponding belief) had caused a different action altogether. (See the essays by Clarke and Kane in this volume.)

Agent causationists typically hold that this is not enough for freedom, or at least for the sort of freedom that can directly ground ascriptions of responsibility. And what this causal indeterminist scenario lacks is precisely the agent's directly controlling the outcome. The agent's internal states have objective tendencies of some determinate measure to cause

certain outcomes. While this provides an opening in which the agent might freely select one option from a plurality of real alternatives, it fails to introduce a causal capacity that fills it. And what better here than it's being the agent himself that causes the particular action that is to be performed?

Such is the intuitive pull in favor of an agent causal account. But pursuing this suggestion leads to a number of theoretical questions: What kind of thing is the 'agent'? What precisely is it that the agent causes? How are agent causation and event causation related? How do the agent's reasons figure into the equation? (Surely they must somehow govern instances of agent causation - but how?) And under what circumstances might agent causation occur? Unsurprisingly, agent causationists (AC) have answered these questions in different ways.

### Some Ontological Requirements

All AC theorists require that we think of agents as things which endure through time, such that they are wholly present at each moment of their existence. This is in contrast to a now popular 'temporal parts' ontology, on which things are supposed to be composed of temporal parts, in much the way that they have spatial parts. According to this view, just as my left foot is but a spatial part of me (and when we say that I exist there, we actually mean that a part of me is there, that I overlap that region), so also the present stage of my existence is a temporal slice of my whole being, a component of the four-dimensional object spanning some eighty (?) years that, speaking tenselessly, is me.<sup>1</sup> Clearly, such a temporally extended object is not suited to play the role of an agent cause of ever so many particular episodes in its own life. But nor are any of the momentary stages suitable, as these are not distinct from total states of the object at a particular time, and agent causation is supposed to be different from causation by states or events within the agent. Hence, there is nowhere to 'put' agent causation in the temporal parts theorist's ontology.

More is required, however, than a rejection of the doctrine of temporal parts. For suppose a contrary view ("presentism") on which all that exists simpliciter, exists at the present moment. Some have supposed, consistent with this, that the general category of objects or substances is somehow reducible to that of events, conceived as localized property instantiations. Again, if this were so, the notion of a distinctively personal causality could not be made out, since its distinctiveness from event causation rests in part on the assumption that agents belong to a distinct ontological category (that of substance).

Finally, even on views that admit enduring substances, many suppose a kind of reductionism concerning composite substances. The being and activity of such composites, they aver, is wholly constituted by the being and activity of their fundamental constituents. If I am a composite biological organism, my activity now - I here speak of the concrete 'token' process embedded in the world, as opposed to any abstract type to which this process belongs - is simply and entirely a structured, logical consequence of the mechanistic activity of my present constituents. And so, once more, there should be no room for a distinctive kind of agency exhibited by me and other persons. So the agent causationist requires an ontology on which persons are enduring, ineliminable substances that are in some robust sense more than the sum of the constituents of their bodies.

#### Does Agent Causation Require Substance Dualism?

It has often been supposed that the only way the requirement that agents not be mere composites can be satisfied is for agents to be wholly distinct substances from their bodies. However, this is not typically argued with any care. Chisholm did hold that agents were simple substances, but the reason he gave for this had to do with his very demanding view on what it takes for any object to endure through time.<sup>2</sup> He was a 'mereological essentialist,' believing that objects have all of their parts essentially. Given that human bodies are like every other composite physical objects in losing parts over time, this implies that an individual human body does not exist for a substantial length of time—or if it does, it is

only as a 'scattered object' that cannot be identified with a human person, much less with someone having an irreducible agent-causal capacity. By contrast, Richard Taylor (1966: 134-38) is quite emphatic that human persons are simply living animals, having no immaterial parts, and Randolph Clarke suggests the same (1993: 201, n.14).

Is there a special reason for the agent causationist to be a dualist? Clearly, agent causal power and its exercise cannot be constituted by underlying event-causal processes, on pain of giving up the claim that it is an ontologically irreducible power. (John Bishop 1983, 1986 held for a time that agent causation was conceptually, but not ontologically, irreducible to event-causal processes. But this would render agent causation useless for the purpose of solving the metaphysical problem of freedom.) It may be enough, though, that one suppose that agent causal power and its allied properties are ontologically emergent, while still being powers and properties of the biological organism. That is, one might embrace a strong form of property dualism, consistent with substance monism. Note that this requires a metaphysical, not merely epistemological, understanding of emergence, and so something rather more ambitious than what is contemplated when the term "emergence" is used in some contemporary theories of mind in philosophy and cognitive science.

O'Connor (2000: Ch.6) defends just such a position. On that account, a state of an object is emergent if it instantiates one or more simple, or nonstructural, properties and is a causal consequence of the object's exhibiting some general type of complex configuration (whose complexity will probably be a feature of both its intrinsic and functional structure). A property is 'nonstructural' just in case its instantiation does not even partly consist in the instantiation of a plurality of more basic properties. An emergent state is a 'causal consequence' of the object's complex configuration in the following way: in addition to having a locally determinative influence in the manner characterized by physical science, fundamental particles or systems also naturally tend (in any context) toward the generation of such an emergent state. But their doing so is not discernible in contexts not exhibiting the requisite macro-complexity, because each such tending on its own is 'incomplete.' It takes

the right threshold degree of complexity for those tendencies, present in each micro-particle, to jointly achieve their characteristic effect, which is the generation of a specific type of holistic state.

Note that agent causation would require the emergence of a very different sort of capacity altogether, one that is distinctive in kind (a fundamentally intentional form of causality) and whose indeterminacy may well not be characterizable in a manner suitable to any quantum indeterminacy. As Jan Cover and John O'Leary-Hawthorne (1996) note, in criticizing this emergentist approach to an agent causal theory, this forces us to posit a sharp break somewhere in the seemingly smooth difference of degree among organisms of increasing complexity. They also suggest that such a theorist is going to have a difficult time saying just what exactly the agent who is doing the causing is. The official answer is the whole living organism, but one might suppose that this is too inspecific. For presumably the agent causal capacity emerges from the right kind of neural structure. (A brain in a vat would have this capacity, too, one assumes.) But perhaps this is too general still. Might it not strictly be some part of the brain? Or perhaps the capacity is associated with different particular parts in different particular circumstances? They worry that one's answer will either attribute (implausibly) some degree of arbitrariness to the basic workings of nature or will require us to suppose (again implausibly) that what kind of thing is the subject of agent causal power is determined extrinsically. (It's the whole organism where there is one, but a brain where it is envatted and disembodied.) Along similar lines, William Hasker (1999) contends that the kind of unity of the agent that is presupposed by the thesis of agent causation directly entails that the agent is a mereological simple.

I'll not try to sort through possible replies to this last worry, involving as it does difficult questions about the metaphysics of composites. (Can there be an individual who is composed of many things yet whose individuality is not a function of - not grounded in - the individuality of those parts, and who indeed can persist in the face of complete changeover of those parts over time? Would the holistic emergent properties of systems as

characterized above suffice for such distinctive individuality? How would one distinguish this kind of emergent composite individuality from an emergent substance dualism, on which the bodily system is wholly non-overlapping the mental substance?) But I note that the other worries, concerning the strong assumptions one must make about the dramatic difference emergence can make is likely to be faced by the substance dualist as well. For it is plausible that if substance dualism is true, it will involve the emergence of mental substances.

Suppose a substance dualist picks up the argument at this point and claims that we have philosophical motivation to want this even stronger kind of emergence: Agents seem too ontologically superficial on the emergent-capacity-only view. On that view, there coming to be exercises of free choices by highly complex systems of matter was foreordained long ago, a direct product of the microphysical fabric of things.<sup>3</sup>

Why not go for the whole hog, then, and embrace an emergentist form of substance dualism, on which agents themselves are radically distinct things? Probably the main objection to doing so is that the causal generation of mind stuff amounts to creation ex nihilo. What some refuse to contemplate even for God is now being contemplated for bits of matter! Note that the kind of 'tending towards the generation of an emergent property or capacity' ascribed to fundamental particles on the more modest emergentist picture is not different in kind from garden variety tendencies. It is a tendency to qualify a system in a certain way - to induce a change in the system's properties. The substance emergence view, by contrast, involves a tendency to generate new stuff. Hasker (1999) denies that this amounts to creation ex nihilo, but his reason is unclear. It may be that, as a theist, he regards the exercise of causal power by any created thing as inherently dependent on God's constant activity of sustaining things in existence. But it would seem that for this to suffice to ward off the charge, one will have to further suppose that God directly plays a further, ineliminable role in the very producing of the emergent substance, so that while it is creation ex nihilo, it is not (entirely) an accomplishment of the physical system.

Others will question the sharpness of the contrast between substance and property emergence by rejecting traditional views of the ontology of substances and their properties on which the contrast is based. The dualist Karl Popper, for example, would not see a sharp difference between the two forms of emergence. If the "property emergence" view posits rich and enduring psychological structures of an emergent sort—as argued in Eccles and Popper (1977)—then, says Popper, one has thereby described a self or person that is distinct from the physical organism.

### What Is the Relation Between Agent And Event Causation?

I now turn to the idea of causation at work in various agent causationist accounts and the consequent similarity or dissimilarity of agent and event causation. Agent causationists universally reject theories that purport to reductively analyze the concept of causation to noncausal notions, such as certain patterns of actual similarity among event types, as on the traditional Humean analysis, or of counterfactual similarity, as on David Lewis's more recent view (1986b). This rejection is unsurprising, since agent causation, understood as a kind of control functioning more or less independently of the agent's dispositional states, clearly cannot be understood in any such terms. Thomas Reid (1788) and George Berkeley (1710 and 1713), the two most prominent defenders of agent causal theories in early modern philosophy, went so far as to hold that agent causation is the only form of causation properly so-called. The regular patterns exhibited in our experience among sensible objects are directly produced by God, the supreme agent cause. (For a fine exposition of Thomas Reid's views on agent and event causation that situates Reid's views alongside more recent accounts, see William Rowe (1991). Also see O'Connor (1994 and 2000: Ch.3).)

Contemporary agent causationists have instead held that agent and event causation are equally basic, related features in the natural order of things, although Chisholm and Taylor betray a lingering tendency to think of event causation as ontologically second-class.<sup>4</sup> In emphasizing differences between event and agent causation, they naturally opened

themselves to the charge that the sole similarity was in the term "causation." They were far from hostile to this - Taylor (1966: 262) says as much - but their critics saw this as casting suspicion on the idea of agent, not event, causation.

In various writings (e.g., 1971 and 1979), Roderick Chisholm contended that the correlated notions of physical necessity and law of nature are primitives in terms of which event causation is to be understood. Roughly speaking, an event A causally contributes to event B just in case it is part of a minimally sufficient condition for B or is essential to preventing a sufficient condition for not-E. (This last case allows for indeterministic event causation. I note that in his various writings on this matter (e.g., 1966, 1976a and 1976b), Chisholm did not always include such a condition.) By contrast, the core agent causal event, which Chisholm termed an "undertaking (or endeavoring) to make [state of affairs] A happen," is understood apart from the concept of law of nature as an essentially intentional form of direct control by an agent. (1967: 413) So in the event causal case, causality is reductively definable in terms of a modal concept (albeit a primitive one, distinct from the fundamental notion of absolute necessity), whereas in the agent causal case, it is a primitive intentional concept, and one which is intended to carry a primitive causal sense to boot. (See 1967: 413, n.6.)

Richard Taylor (1966) also supposed that a primitive, logically contingent form of necessity figures into our understanding of event causation. On his view, event or condition A was the total cause of event B only if each was a 'necessary and sufficient condition' for the other in the circumstances. But in contrast to Chisholm's reductive analysis, he added the requirement that A made B happen by virtue of its power to do so (p.38). This forges a link between agent and event causation: both are manifestations of primitive powers or capacities. The difference between them consists in the different types of entities that are causes and the modal feature that event causes, but not agent causes, necessitate their effects. Indeterministic event causation is a conceptual impossibility for Taylor; if our best fundamental physical theories posit merely statistical regularities, then they imply a lack of

causality. But the existence of the agent with his distinctive capacities does not necessitate any particular effects. Taylor does allow that circumstances might necessitate the causally complex event of the agent's causing some event B, in which case the agent would not be acting freely. (He denies, however, that this is typically the case.) This idea is problematic, as it implies that a condition might directly make happen the obtaining of a causal relation. (See O'Connor 1996, 2000.) But it is quite independent of the rest of what Taylor says and is not of central interest here.

More recently, the present author has defended a conception of the relation of agent and event causation that is a hybrid of the above accounts. (See O'Connor 1995 and 2000.) With Chisholm, I hold that agent causation alone is essentially intentional and purposive. The fundamental locution I employ is "an agent's causing an intention for a reason."<sup>5</sup> (Taylor, by contrast, holds that an agent might cause an event for no reason at all.) But I also agree with Taylor that event causation, too, involves the exercise of primitive capacities, though I deny that all such capacities must be structured in such a way that event causes invariably necessitate their effects. Deterministic propensities are but the limiting cases of probabilistic tendencies. In saying this, we should not suppose with some (e.g., Paul Humphreys 1989) that an indeterministic tendency is something that merely structures what is a 'chance' outcome. Whatever happens is made to happen by its cause. That the cause operated indeterministically implies only that it might not have produced that outcome—it had a positive tendency in the total circumstances towards more than one type of outcome.

A final and markedly different understanding of agent and event causation and their relationship has recently been proposed by Randolph Clarke. Clarke's point of departure is the novel analysis of event causation proposed independently by Fred Dretske (1977), David Armstrong (1983), and Michael Tooley (1977, 1987). This view eschews primitive dispositionality in favor of a primitive type of relation. In basic outline, the view identifies laws of nature with certain primitive, contingent, and second-order relations among universals, ones that are specified as satisfying certain theoretical requirements associated

with our concept of scientific law. The event causal relation, conceived as a type, is a special subset of these and is instanced between first-order events.

Clarke suggests that the very relation of causation that is thus theoretically identified within the domain of complex universals also holds between agents and their actions in instances of freely performed action. (The agent causationist can ride piggy back on the proposed understanding of the causal relation provided by its role in structuring patterns among events.) The sole differences between event and agent causation are the causal relations and the contingent fact that event causings are structured by probabilistic or deterministic laws. (But see below for more on Clarke's view on agent causation and causal laws.)

Now the second-order relation view of the causal relation is not without its problems. Bas van Fraassen (1988: Ch.5) has challenged the relevance of the posited relation to explaining any particular causal sequence: How does there being a second-order relation among pairs of complex universals explain why particular event A brought about particular event B? After all, not all properties of types, including relational properties, carry over to their tokens. Van Fraassen calls this 'the Inference Problem': the problem of explaining why we're entitled to infer from the posited second-order relation among pairs of universals that this particular event or state of affairs, instantiating the first member of one of the pairs, will cause an instantiation of the corresponding member.

My purpose in drawing attention to this problem for the general view is to note a certain direction within the responses of the view's proponents. David Armstrong (1997) has come to hold that causation is manifested in our world as simply and solely a relation among types of states of affairs. (So when I experience the causal force, say, that is exerted on my toe by a heavy object, what I am experiencing is nothing particular, but rather causation in general, or nomicity.) And Michael Tooley's 'speculative proposal' in response to the Inference Problem is to posit unusual features in the mereology of transcendent universals. If it is a law of nature that all things having property P have property Q, then, he says, we might suppose that P "exists only as part of the conjunctive universal, P and Q"

(1987: 124). It would then follow that any time P is instantiated, Q is as well, thus grounding the inference on the observation of a P-type event. Both of these responses put pressure on the official view that the second-order relations by means of which we specify the causal relation are only contingently associated with it (as clearly is true when we use the contingent fact about Socrates that he was the teacher of Plato as one means of identifying him).

The present significance of this is that by supposing associations of these kinds to be essential to the causal relation, as I myself think proponents of the general program should, one will also require that agent causation be law governed. Now Clarke suggests that will be so, in any case. We might suppose laws of nature to the effect that the causal relation obtains between agents and certain events only where agents have properties requisite to reflective practical reasoning and the caused events are instances of acting for reasons. Further, it might be a law of nature that whenever agents with such capacities do act on reasons, the causal relation obtains between the agent and the action (though the laws and antecedent circumstances do not imply which action will be so caused.) But if this is so, the explanation for their being a causal relation now between me and my action appears to reside in these general lawlike facts about agents-in-general, not me in particular. That is, it's not clear that I am fundamentally responsible for this result. Bear in mind that on the general causal view in question, there are no primitive, single-case dispositions at work. The significance of calling the posited relation 'causal' instead lies in the higher-order, completely general facts about how this relation structures patterns of property instantiation in the world. One worries, then, that this general view is less hospitable than Clarke supposes to the agent-causalist's motivation to ground ascriptions of fundamental responsibility for certain outcomes to agents.

What Does the Agent Cause?

As a rule, agent causationists are surprisingly vague on the basic question of just what agents cause. Chisholm and Taylor both repudiate choices or volitions as a basic mental category. Taylor says that agents cause their behavior (the whole sequence?) and Chisholm says that in trying to bring about some state of affairs, the agent makes happen some more immediate state of affairs, which he supposes is neurophysiological in character. This neurophysiological event is not identified with any intentional state, though, as we've seen, the agent's causing it is intentional. Finally, while Clarke does speak of agents as making choices, he consistently says that agents cause their actions, and so seems to have a view similar to Taylor's on this score (though they differ on what features warrant characterizing the behavior as an 'action'). C.A. Campbell (1967), by contrast, speaks of the self as directly determining a decision, and Michael Zimmerman (1984) of the agent's directly effectuating a volition.<sup>6</sup> Along similar lines, Donagan (1987) and O'Connor (1995, 2000) hold that the agent causes an immediately executive, or action-triggering, intentional state. (Strictly speaking, the agent's causing such an intentional state is what they term the agent's 'choice', and it is also the agent's basic action, typically constituting the initial segment of more extended event-causal processes resulting from such choices.)

#### IV Agent Causation and Reasons Explanation

C.A. Campbell (1967) held that self causality is conceivable only in the special circumstance where our 'desiring nature' is opposed to our perception of moral duty (p.46). In all other cases, he supposed, our formed character will inevitably result in some particular outcome, even if it is not immediately apparent to us which outcome that will be. But in this special situation of moral temptation, nothing other than the agent or self determines what happens. If one asks why the agent acts as he does on a given occasion - now succumbing to temptation, now doing what he ought - the correct answer, Campbell insists, is that there is no explanation to be had. (Seemingly at odds with this, however, he does allow that some situations of temptation are harder than others, and that consequently we may reasonably

expect resistance to temptation to occur less frequently in such cases.) Though this view is implausibly restrictive of the scope of free action, it does have the advantage of evading the difficult matter facing most agent causationists of how reasons guide and thereby explain the agent's exertion of causal power.

Chisholm suggested that reasons are necessary causal conditions on agent's causing their actions. I am always acting with some purpose, and my desiring to attain that end and having appropriate beliefs about how to do so thereby contribute to my doing so, not by forming part of a sufficient condition for the action, but by their being essential to preventing the occurrence of a sufficient condition for my not causing the action. (Had those factors not obtained, there would have been a sufficient condition for my not causing that particular action.) One problem with this way of understanding the role of reasons is that we can envision cases where my having reason A and my having reason B each guide my performing an action but neither of which is such that, had that state not obtained, I would have been precluded from performing the action. (As should be apparent, this is most directly a problem for Chisholm's modal analysis of event causation.)

Taylor had a quite different view. According to him, explanations in terms of reasons or purposes are entirely different from explanations in terms of causes (p.142).<sup>7</sup> When we recall that Taylor repudiates any intentional event which triggers and guides the completion of the action, it becomes puzzling how the agent's having a purpose guides his causing the action. The action's initial segment will be constituted by enormously complex neuronal events, none of which the agent is consciously aware of. So how does he effortlessly 'get it right,' causing just the right complex sequence for an action that will carry out his purpose? (We cannot suppose, for example, that the agent's having the purpose is a state that is governed by causal laws that map it onto the appropriate outcomes.) Furthermore, it's not clear how Taylor can meet Davidson's famous challenge to non-causal understandings of reasons explanations: Among cases where the agent has more than one reason for performing an action, it's plausible to suppose than in some of them only one reason

actually prompted the action, while in others there was a plurality of factors that did so. In what does this difference consist?

In my (1995), I follow Taylor in construing reasons explanations as non-causal. However, I contend that a satisfactory answer to Davidson's challenge requires (what is independently plausible on introspective grounds) the agent causationist to suppose that agents cause executive states of intention of a particular sort. The content of these intentions is not merely that I perform an action of type  $\emptyset$ , but that I perform an action of type  $\emptyset$  *in order to satisfy desire D* (or prior intention I). If intentions have this rich sort of content, then the difference between acting to satisfy desire D1 and acting to satisfy D2 and acting to satisfy both, will be a function of the content of the intention that I cause to occur. When Davidson asks what accounts for my acting on reason R1 and not R2, given that I was aware of both at the time of acting, the answer will be that we must look to the content of the intention I cause; this will have the form, that I do A for in order to satisfy reason \_\_\_\_\_. In a given case, the blank will be filled by either or both of R1 and R2. In actively deciding which action I will undertake, I am inter alia deciding which reason I am aiming to satisfy.

Note that it is also required that this intention, once generated, causally sustains the completion of the action in an appropriate manner. We thereby avoid the sort of counterexample Alfred Mele (1992) raises against purely non-causal accounts of reasons explanations, such as that of George Wilson (1980), in which there is a causal process independent of the intention that generates the action.

Randolph Clarke objects that since the intention refers to the action, it must be causally or explanatorily posterior to the action, and so the account absurdly implies that what explains the action depends on the outcome of something posterior to it in this way. But this rests on a misunderstanding. On the proposed account, the agent directly causes (the coming to be of) a state of intention. This constitutes the 'core' action. The intention refers not to some independent process which is the action, but an action sequence of which it is the initiating segment.

Finally, Galen Strawson (1986: Ch.2; 1994) has objected to indeterministic theories of free action generally that they (unwittingly) entail an infinite regress of choices corresponding to every indeterministic choice of a course of action. Since how one acts is a result of, or explained by, 'how one is, mentally speaking' (M), for one to be responsible for that choice one must be responsible for M. To be responsible for M, one must have chosen to be M itself — and that not blindly, but deliberately, in accordance with some reasons R1. But for that choice to be a responsible one, one must have chosen to be such as to be moved by R1, requiring some further reasons R2 for such a choice. And so on, ad infinitum. Free choice requires an impossible regress of choices to be the way one is in making choices.

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

But let us consider what the agent causationist might say in reply to Strawson, on the agent-causal account I sketched above. Aware of certain reasons pro and con, I cause an action-initiating intention to A for reason R1. This is explained by my having been aware of reason R1 while deliberating and as I completed the action. I did not directly choose to be in a state of being aware of and motivated by R1. I simply found myself in that state, among others, and proceeded to deliberate. Being in that sum total of rational states circumscribed

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

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

Surely one must concede in response that responsibility for 'shaping who I am' and for the choices that ensue from this comes in degrees and, indeed, can only sensibly be measured within a limited scope of possibilities. We cannot hold Billy responsible for failing to consider an option entirely outside the range of his experience. And his responsibility for passing by options which are within the range of his experience but which

he has had precious little opportunity to consider as attractive is attenuated. In concrete cases, given limited information, we hazard rough guesses on these matters. When we are confronted with an individual who quite deliberately and unhesitatingly makes a grossly immoral choice - indeed, who seems not to even consider the obvious moral alternative - the question one needs to ask is this: was there a point earlier in her life when paths was open to her (ones for which at each step of the way she had some significant motivation to pursue, and which she recognized as having moral significance) such that had she taken them she would now be such as to see the force of the moral considerations at hand? How 'difficult' would it have been for her to pursue such a path? Our guesses about such matters are exceedingly rough, and rely on the assumption that a certain measure of rough moral sensitivity is had by most mature individuals. Absent compelling information to the contrary, then, we deem it appropriate to hold individuals responsible for their own moral indifference.

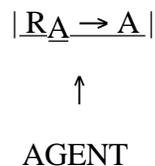
Perhaps the important point to emphasize here is that the agent causationist can consistently allow that there are individuals whose basic choice-making capacities are just like ours, but who lack sensitivity to a variety of forms of basic human decency through no fault of their own. Furthermore, he can also accept (what may also be behind Strawson's argument) that perfect responsibility for one's choices and character is not just contingently lacking in human beings, but is an impossible notion: it would require perhaps perfect indifference at the outset, or at least an openness to all possible courses of action. The coherence of that idea is very doubtful.

#### V Mechanism and Agent Causation: Narrowing the Gap?

In the preceding discussion, I frequently adverted to the boringly familiar thought that our motivation to pursue courses of action varies considerably. Evident as this is, one might well suppose that some agent causationists have not sufficiently taken this fact to heart in their accounts of the way reasons explain actions. Taylor (1966) and O'Connor (1995), for

example, both lack any account of strength of desire and settled preferences. This doubtless figures into Taylor's rejection of any causal role for reasons (see 1966: 250) and in the very attenuated role O'Connor (1995) assigns, in which my now having a reason to A is a necessary causal condition on my now causing the intention to A.

Randolph Clarke (1993, 1996) develops his account of agent causation with the explicit goal of remedying this defect in the traditional picture. In the early version, when an agent acts freely, her coming to have reasons to so act ( $R_A$ ) indeterministically cause her action (A). The agent figures into the picture by causing, not the action simpliciter, but the action's being done for those reasons. If we let " $\rightarrow$ " stand for the causal relation, we may diagram Clarke's basic picture thus:



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

In his (1996), Clarke maintains instead that the agent and the agent's indeterministic state of having reason  $R_A$  jointly produce A. The agent's causal capacity consists in the ability to make effective an indeterministic propensity of one's reason to bring about A, not by directly producing a causal relation between two events, but in the sense of 'acting

alongside' or bolstering the tendency (whether it be of a low or high probability measure), ensuring that it will achieve its characteristic effect.

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

Clarke does not elaborate on this hegemony of the agent's causality over the tendencies conferred by one's reasons, but it seems to me to point to a promising way to achieve the desired integration. In my (2000), I suggest that my coming to recognize a reason to act induces or elevates an objective propensity for me to initiate the behavior. That is to say, agent causation is a probabilistically structured capacity. It will be structured not only by tendency-conferring states of having reasons to act in specific ways but also by more enduring states of character, involving relatively fixed dispositions and long-standing general intentions and purposes around which my life has come to be organized. I am the sole causal factor directly generating my intention to A (not a co-cause along with my reasons, as on Clarke's view), but my doing what I do is shaped - causally shaped - by my total motivational state.

Embracing this causal propensity account of the relative strength of reasons—and perhaps even supposing it to be required to make sense of the very idea that reasons in general motivate actions—need not lead one to abandon the non-causal link suggested above between actions and the reasons which explain them. For the mere fact that a reason I had gave me some tendency in this sense to act as I did does not explain my action. Maybe I did not act for that reason, despite my recognition that it was a relevant reason. (I wanted to

save Charles pain, but that's not, even in part, why I pulled the plug.) Within the framework of an agent-causal account, this is still to be determined by the content of the agent's intention in acting.

Or so it seems. Randolph Clarke (1996: 47, n.37) challenged the original content-of-the-intention approach to reasons explanation in a way that might be transposed as follows to apply to the present, more complicated picture: The account appears to allow that the purpose-identifying content of my intention might not refer to a motivational state conferring a very high propensity for me to cause the type of action I did cause. But surely, he objects, such a state must figure into the true explanation of my acting as I did. I can't declare by fiat in my intention that guides my action what had been my motivation in so acting. My reply is that this last is surely so, but I do form the intention, and this is not an accompanying declaration of some kind, possibly mistaken, but the initiating core of what I do, and its content guides the completion of the action. The fact that a particular reason gave me a strong propensity to act in a particular way certainly explains in part my considering that possibility while deliberating; but if I cease to consider this factor when I cause my intention to act in this way—an odd scenario, to be sure, but seemingly possible—and actively form an action with the intentional content of satisfying another reason, then that alone explains why I acted just as I did.

(Are we then saying as a general thesis about action explanations that reasons that are referred to in the content of any agent's intentions automatically explain? Surely we can conceive an external manipulation of one's choice formation such that one is caused to do A with the intention of doing A for reason R, although R is in no way explanatory. But the agent causationist need make no such general claim. The agent-causal origin of the intention is a necessary ingredient of this sort of explanatory framework. It makes the coming to be of the content of the intention something I am directly responsible for and (if the argument of O'Connor 1996 and 2000 is correct) it precludes an independent causal explanation.)

## Bibliography

- Armstrong, David. 1983. What Is a Law of Nature?. Cambridge: Cambridge University Press.
- 1997. A World of States of Affairs. Cambridge: Cambridge University Press.
- Berkeley, George. 1710. Principles of Human Knowledge.
- 1713. Three Dialogues Between Hylas and Philonous.
- Bishop, John. 1983. "Agent-causation." Mind 92: 61-79.
- 1986. "Is Agent-Causality a Conceptual Primitive?" Synthese 67: 225-47.
- 1989. Natural Agency. Cambridge: Cambridge University Press.
- Campbell, C.A. 1967. In Defence of Free Will & other essays. London: Allen & Unwin.
- Chisholm, Roderick M. 1966. "Freedom and Action." K. Lehrer, ed. Freedom and Determinism. New York: Random House.
- 1967. "He Could Have Done Otherwise." Journal of Philosophy 64: 409-18.
- 1971. "Reflections on Human Agency." Idealistic Studies 1: 33-46.
- 1976a. "The Agent as Cause." M. Brand and D. Walton, eds. Action Theory. Dordrecht: D.Reidel.
- 1976b. Person and Object. LaSalle: Open Court.
- 1979. "Objects and Persons: Revisions and Replies." E. Sosa, ed. Essays on the Philosophy of Roderick Chisholm. Amsterdam: Rodopi.
- Clarke, Randolph. 1993. "Toward a Credible Agent-Causal Account of Free Will." Noûs 27: 191-203.
- 1996. "Agent Causation and Event Causation in the Production of Free Action." Philosophical Topics 24 (Fall): 19-48.
- 1997. "On the Possibility of Rational Free Action." Philosophical Studies 88: 37-57.
- Cover, Jan and O'Leary-Hawthorne, John. 1996. "Free Agency and Materialism." J. Jordan and D. Howard-Snyder, eds. Faith, Freedom, and Rationality. London: Rowman and Littlefield.

- Davidson, Donald. 1963 [1980]. "Actions, Reasons, and Causes." Essays on Actions and Events. Oxford: Clarendon Press.
- Donagan, Alan. 1987. Choice. London: Routledge and Kegan Paul.
- Dretske, Fred. 1977. "Laws of Nature." Philosophy of Science 44: 248-68.
- Eccles, John and Popper, Karl. 1977. The Self and Its Brain. New York: Springer-Verlag.
- Hasker, William. 1999. The Emergent Self. Ithaca: Cornell University Press.
- Heller, Mark. 1990. The Ontology of Physical Objects: Four-Dimensional Hunks of Matter. Cambridge: Cambridge University Press.
- Humphreys, Paul. 1989. The Chances of Explanation. Princeton: Princeton University Press.
- Kim, Jaegwon. 1998. Mind in a Physical World. Cambridge, MA: MIT Press.
- Lewis, David. 1986a. On the Plurality of Worlds. Oxford: Blackwell.
- 1986b. "Causation." Philosophical Papers, vol. 2. Oxford: Oxford University Press.
- Malcolm, Norman. 1968. "The Conceivability of Mechanism." Philosophical Review 77: 45-72.
- McCann, Hugh. 1998. The Works of Agency. Ithaca: Cornell University Press.
- Mele, Alfred. 1992. Springs of Action. New York: Oxford University Press.
- 1995. Autonomous Agents. New York: Oxford University Press.
- O'Connor, Timothy. 1993. "Indeterminism and Free Agency: Three Recent Views." Philosophy and Phenomenological Research 53: 499-526.
- 1994. "Thomas Reid on Free Agency." Journal of the History of Philosophy 32: 605-22.
- 1995. "Agent Causation." T. O'Connor, ed. Agents, Causes, and Events: Essays on Indeterminism and Free Will. New York: Oxford University Press.
- 1996. "Why Agent Causation?" Philosophical Topics 24 (Fall): 143-158.
- 2000. Persons and Causes: The Metaphysics of Free Will. New York: Oxford University Press.

- Reid, Thomas. 1788. Essays on the Active Powers of Man.
- Rowe, William. 1991. Thomas Reid on Freedom and Morality. Ithaca: Cornell University Press.
- Strawson, Galen. 1986. Freedom and Belief. Oxford: Clarendon Press.
- 1994. "The Impossibility of Moral Responsibility." Philosophical Studies 75: 5-24.
- Swinburne, Richard. 1986. The Evolution of the Soul. Oxford: Clarendon Press.
- Taylor, Richard. 1966. Action and Purpose. Englewood Cliffs: Prentice Hall.
- Thorp, John. 1980. Freewill: A Defence Against Neurophysiological Determinism. London: Routledge and Kegan Paul.
- Tooley, Michael. 1977. "The Nature of Law." Canadian Journal of Philosophy 7: 667-98.
- 1987. Causation: A Realist Approach. Oxford: Clarendon Press.
- Trusted, Jennifer. 1984. Free Will and Responsibility. Oxford: Oxford University Press.
- Van Fraassen, Bas. 1988. Laws and Symmetry. Oxford: Clarendon Press.
- Van Inwagen, Peter. 1990. "Four-Dimensional Objects." Noûs 24: 245-256.
- Vesey, Godfrey and Flew, Antony. 1987. Agency and Necessity. Oxford: Blackwell.
- Wilson, George. 1980. The Intentionality of Human Action. Amsterdam: North-Holland.
- Zimmerman, Michael. 1984. An Essay on Human Action. New York: Peter Lang.

---

<sup>1</sup> For accounts and defenses of temporal parts ontology, see David Lewis (1986a) and Mark Heller (1990). The view is criticized by Peter van Inwagen (1990).

<sup>2</sup> Richard Swinburne (1986) may be advocating agent causation with his notion of active 'purposings,' although he is not explicit on this point. Purposings are likened to volitions in some other accounts, though Swinburne emphasizes that purposings (1) may merely consist in allowing certain actions to occur, rather than consciously choosing them and (2) are intrinsically active and having content that points to the result sought. But these claims are accepted, for example, by Hugh McCann (1998) - discussed in Clarke's contribution to this

---

volume - and McCann is not an agent causationist. Swinburne is also a dualist, but like Chisholm, his motivations concern problems with identity over time rather than reasons specific to freedom of action.

<sup>3</sup> In correspondence and unpublished work, Peter Unger has advanced this argument in favor of substance dualism over softer varieties.

<sup>4</sup> Jennifer Trusted (1984) appears to defend an agent-causal account of human agency, the capacity for which emerges from event-causal physical activity. But I am not at all confident in interpreting her final view, developed over the course of a wide-ranging discussion, much of which exposites the views of others.

<sup>5</sup> William Rowe (1991), in interpreting Thomas Reid's view, speaks of an agent's "exerting active power." But he wishes to contrast this with an allegedly mysterious view on which there is an "irreducible relation" between the agent and his willing, a view he sees in Chisholm and Taylor. (See pp.156-7.) I myself am unclear on what an exertion of active power is, if it does not consist precisely of an event in which an agent causes some event, and this is how I understand Reid himself. Perhaps all that Rowe's remarks on this point come to is an insistence that an agent's exerting active power is something he does, a point on which Rowe may take Chisholm and Taylor to be unclear. If this is correct, then my earlier (1994) understanding of Rowe's account, on which Reid is read as a noncausal theorist of the sort discussed in Clarke's contribution to this volume, is mistaken and the consequent criticisms misplaced.

<sup>6</sup> John Thorp (1980: 102) writes: "Now presumably we shall want to say that the agent's causing the event is also an event. We seem then to have two events, the decision which is an alteration in the agent, and the agent's causing that alteration. At once there looms a vicious regress. It can be forestalled only by saying that these apparently two events, the decision and the agent's causing the decision to itself, are in fact one and the same.... We do not require that an event be the same as its cause, but that an event be the same event as its being

---

caused." It is not clear to me from the wider text whether this is a (misleading) way of saying that an agent's causing an event is not itself an event or whether he is effectively reducing agent causation to simple indeterminism.

<sup>7</sup> Indeed, he held the puzzling thesis that there might be completely independent purposive and causal explanations for the very same action (p.144).