

# Agent-Causal Power

TIMOTHY O'CONNOR

## Abstract

*Some philosophers have thought that an adequate account of free will must be centered on a notion of 'agent causation', an ontologically primitive type of causation that is uniquely manifested by (some possible) persons – a causality that is inherently goal-directed and nondeterministic. While having noted historical adherents, this view is widely dismissed as unintelligible. In this essay, I seek to reverse this turn of events by explicating agent-causal power against the backdrop of a more general 'causal powers' account of causation.*

Our universe is populated, at bottom, by a vast number of partless particulars of a few basic kinds. Each of these particulars instantiates some of a small range of primitive qualities and stands in primitive relations with other particulars. We may conceive of qualities as immanent universals that are numerically identical in their instances, along the lines championed by David Armstrong (1978), or as 'tropes' that are non-identical but exactly resembling in their instances, following C. B.

Martin (1997). (Nothing of what follows hangs on which alternative one prefers.) In calling these qualities “primitive,” we are contending that they are real existents whose being does not consist, even in part, in the existence or instantiation of other entities. The nature of these qualities is irreducibly dispositional: they are tendencies to interact with other qualities in producing some effect, or some range of possible effects. (Or, in certain cases – as we shall consider below – they may instead confer upon their possessor a tendency towards, or power to produce, some such effects.) It *may* be that the dispositional profile of a quality does not exhaust its nature, so that it has a further ‘qualitative character,’ or quiddity. Here again, we may be neutral on this question for present purposes.<sup>1</sup>

The formal character of a feature’s dispositionality may be variable. For a couple of centuries after Newton, it was usual to conceive dispositions deterministically: given the right circumstances, causes strictly necessitate their effects. But fundamental physics since the early twentieth century encourages the thought that dispositions may be probabilistic, such that there are objective probabilities less than one that a cause will produce its characteristic effect on a given occasion. On

---

<sup>1</sup> For discussion, see especially Molnar 2003, Heil and Martin 1999, and Hawthorne 2001.

this view, deterministic propensities are simply a limiting case of probabilistic ones.

What is more, it seems possible to conceive pure, unstructured tendencies, ones that are nondeterministic and yet have no particular probability of being manifested on a given occasion.

The qualities of composite particulars (if we do not here embrace eliminativism) are typically structural, consisting in the instantiation of qualities of and relations between the composite's fundamental parts.<sup>2</sup> Very often, the terms we use to refer to features of composites do not pick out such structural properties, since they are insensitive to minor variation in the composite's exact underlying state. (A water molecule's being in a structured arrangement of hydrogen and oxygen atoms persists through small-scale changes in its microphysical composition and state. Of course, there are also much more dramatic cases of underlying variation across instances consistent with applicability of the same macroscopic term, with

---

<sup>2</sup> David Armstrong has been the key contemporary figure in developing the idea of structural properties as part of a 'sparse' ontology. Nonetheless, he waffles a bit in his understanding of their reducibility. He speaks of them as something subtly 'extra' – distinct from the underlying instances of properties and relations but strongly supervening upon them (Armstrong 1997: 37). But this is hard to square with his adamant contention that they are an "ontological free lunch." (See the discussion on 34–45.)

Toby 1/15/08 2:45 PM

**Deleted:** ontological free lunch.

functional terms providing obvious examples.) Here, we should say that the *concept* is satisfied in different cases in virtue of the instantiation of different structural qualities. Despite superficial ways of speaking, there is no distinctive causal power attached to the satisfaction of such a concept, as that would require the causal powers theorist to accept an objectionable form of double counting of causes, one for the macroscopic structural property (itself nothing over and above the instantiation of microscopic properties and relations) and one for the putative multiply-realized functional ‘property.’

Such are some central elements of a causal powers metaphysic. There are many disputes of detail among its adherents – ~~to the disputed matters already noted,~~ we should add the nature of primitive external relations and the substantiality of space or spacetime.<sup>3</sup> Such disputes aside, the causal powers metaphysics stands opposed both to the neo-Humean vision that David Lewis (1983b, 1986c) popularized and

Toby 1/15/08 2:45 PM

**Deleted:** in addition to those already noted

---

<sup>3</sup> On the nature of fundamental external relations, see Molnar 2003: chap. 10; Ellis 2001; and the skeptical discussion in Armstrong’s contribution to Armstrong, Martin, and Place 1996.

the second-order Humean metaphysics of causal realism without causal powers defended in recent years by David Armstrong (1997) and Michael Tooley (1997).<sup>4</sup>

In what follows, I shall presuppose the ecumenical core of the causal powers metaphysics. The argument of this paper concerns what may appear at first to be a wholly unrelated matter, the metaphysics of free will. However, an adequate account of freedom requires, in my judgment, a notion of a distinctive variety of causal power, one which tradition dubs “agent-causal power.” I will first develop this notion and clarify its relationship to other notions. I will then respond to a number of objections either to the possibility of a power so explicated or to its sufficiency for grounding an adequate account of human freedom.

Toby 1/15/08 2:45 PM  
Deleted: try to

---

<sup>4</sup> I characterize the Armstrong–Tooley view as a gratuitously complicated kind of Humeanism because, despite their intentions, second-order relational structure to the world (their  $N(F, G)$  necessitation relation) cannot *explain* first-order facts (instances of  $G$  regularly following instances of  $F$ ) since the second-order relations presuppose the supposed explananda. Were God to construct a Tooley-Armstrong world, determining the distribution of first-order  $F$  and  $G$  facts must precede decisions about  $N(F, G)$  facts.

## 1 The Problem of Freedom and Agent-Causal Power

Central to the notion of metaphysical freedom of action, or free will, is the thought that what I do freely is something that is “up to me,” as Aristotle says. A familiar, disputed line of reasoning, which turns upon the principle that where certain truths that are not up to an agent logically (and so unavoidably) entail some further truth, the latter truth is itself not up to the agent, concludes that nothing of what we do would be up to us were our actions to be embedded within a strictly deterministic universe.<sup>5</sup> But if determinism would threaten our freedom, so would the complete absence of any intelligible causation of our actions. So one might steer a middle course by supposing that the world is governed throughout by unfolding causal processes, just not deterministic ones. In particular, one might suppose that our free actions are caused by our own reasons-bearing states at the time of the action. An action’s being “up to me”, on this view, consists in the nondeterministic causal efficacy of certain of my reasons: I might have performed a different action in identical circumstances because there was a non-zero (and perhaps pronounced) chance that other reasons-bearing states had been efficacious

Toby 1/15/08 2:45 PM

Deleted: appropriate

---

<sup>5</sup> Peter van Inwagen has heavily influenced discussion of this argument over the last thirty years by his careful formulation of the argument. See van Inwagen 1983 and O’Connor 2000 for a friendly amendment.

Toby 1/15/08 2:45 PM

Deleted: three

in producing the action which they indicated.<sup>6</sup> However, according to many critics (myself among them), indeterminist event-causal approaches falter just here, in the fact that the free control they posit is secured by an *absence*, a removal of a condition (causal determination) suggested by the manifestly inadequate varieties of compatibilism. If there is no means by which I can take advantage of this looser connectivity in the flow of events, its presence can't confer a greater kind of control, one that *inter alia* grounds moral responsibility for the action and its consequences.<sup>7</sup> Given the causal indeterminist view, if I am faced with a choice between selfish and generous courses of action, each of which has some significant chance of being chosen, it would seem to be a matter of luck, good or bad, whichever way I choose, since I have no means directly to settle which of the indeterministic propensities gets manifested.

The familiar considerations just sketched lead certain philosophers to conclude that the kind of control necessary for freedom of action involves an ontologically

---

<sup>6</sup> The most prominent advocate of this view of freedom has been Robert Kane. Kane has developed his view in a number of writings, culminating in *The Significance of Free Will*.

<sup>7</sup> A few preceding sentences are borrowed from O'Connor and Churchill 2004.

Toby 1/15/08 2:45 PM

Deleted: , as it were,

primitive capacity of the agent directly to determine which of several alternative courses of action is realized. In these instances of *agent causation*, the cause of an event is not a state of, or event within, the agent; rather, it is the agent himself, an enduring substance.

I will not defend the claim that any actual agents have such a capacity. The philosophical claim is that agents must exercise such a capacity if they are to have metaphysical freedom, but it is an empirical claim whether human agents do exercise such a capacity (and thus do in fact act freely). Our concern is what such an agent-causal power would be, and, given the analysis, whether any agents *could* possibly have it, *given fairly modest assumptions about the agents and their embeddedness within an environment.*

We begin by considering what the notion of agent causation requires us to assume concerning the nature of the agent having such a power. It is commonly supposed that the notion requires a commitment to agents as partless and nonphysical entities. A radically distinct kind of power, the thought goes, requires a radically distinct kind of substance. And this seems further encouraged by my earlier contention that the causal powers metaphysics pushes one to regard typical 'high-level' features as structural properties wholly composed of microphysical properties arranged in a certain way.

But there is more to be said. While the tidiness of substance dualism has its appeal, it is in fact optional for the metaphysician who believes that human beings have ontologically fundamental powers (whether of freedom or consciousness or intentionality). For we may suppose that such powers are *emergent* in the following sense: (i) They are ontologically basic properties (token-distinct from any structural properties of the organism). (ii) As basic properties, they confer causal powers on the systems that have them, powers that non-redundantly contribute to the system's collective causal power, which is otherwise determined by the aggregations of, and relations between, the properties of the system's microphysical parts. Such non-redundant causal power necessarily means a difference even at the microphysical level of the system's unfolding behavior. (This is not a violation of the laws of particle physics but it is a supplement to them, since it involves the presence of a large-scale property that interacts with the properties of small-scale systems.) In respects (i) and (ii), emergent powers are no less basic ontologically than unit negative charge is taken to be by current physics. However, emergent and microphysical powers differ in that (iii) the appearance of emergent powers is caused (*not* 'realized') by the joint efficacy of the qualities and relations of some of the system's fundamental parts and it persists if and only if the overall system maintains the right kind of hierarchically-organized complexity, a kind which must

be determined empirically but is insensitive to continuous small-scale dynamical changes at the microphysical level.<sup>8</sup>

One cannot give uncontroversial examples of emergent properties. Though there are ever so many macroscopic phenomena that seem to be governed by principles of organization highly insensitive to microphysical dynamics, it remains an open question whether such behavior is nonetheless wholly determined, in the final analysis, by ordinary particle dynamics of microphysical structures in and around the system in question.<sup>9</sup> Given the intractable difficulties of trying to compute values for the extremely large number of particles in any medium-sized system (as well as the compounding error of innumerable applications of approximation techniques used even in measuring small-scale systems), it may well forever be impossible in practice to attempt to directly test for the presence or absence of a

---

<sup>8</sup> Concepts of emergence have a long history – one need only consider Aristotle’s notion of irreducible substantial forms. Their coherence is also a matter of controversy. For an attempt to sort out the different ideas that have carried this label, see O’Connor and Wong 2006. And for a detailed exposition and defense of the notion I rely on in the text, see O’Connor and Wong 2005.

<sup>9</sup> For numerous examples of such phenomena, see Laughlin, Pines, Schmalian, Stojkovic, and Wolynes 2000.

truly (ontologically) emergent feature in a macroscopic system. Furthermore, it is difficult to try to spell out in any detail the impact of such a property using a realistic (even if hypothetical) example, since plausible candidates (e.g., phase state transitions or superconductivity in solid state physics, protein functionality in biology, animal consciousness) would likely involve the simultaneous emergence of multiple, interacting powers. Suffice it to say that if, for example, the multiple powers of a particular protein molecule were emergent, then the unfolding dynamics of that molecule *at a microscopic level* would diverge in specifiable ways from what an ideal particle physicist (lacking computational and precision limitations) would expect by extrapolating from a complete understanding of the dynamics of small-scale particle systems. The nature and degree of divergence would provide a basis for capturing the distinctive contribution of the emergent features of the molecule.

As sketchy as the foregoing has been, we must return to our main topic. I have suggested that we can make sense of the idea of ontologically emergent powers, ones that are at once causally dependent on microphysically-based structural states and yet ontologically primitive, and so apt to confer ontologically primitive causal power. If this is correct, then the fact that agent causal power would be fundamental, or nonderivative, does not imply that the agent that deploys it be anything other than a mature human organism. It is simply an empirical question

whether or not the dispositions of the ultimate particles of our universe include the disposition to causally generate and sustain agent-causal power within suitably organized conscious and intelligent agents.

One important feature of agent-causal power is that it is not directed to any particular effects. Instead, it confers upon an agent a power to cause a certain type of event within the agent: the coming to be of a state of intention to carry out some act, thereby resolving a state of uncertainty about which action to undertake. (For ease of exposition, I shall hereafter speak of “causing an intention,” which is to be understood as shorthand for “causing an event which is the coming to be of a state of intention.”) This power is multi-valent, capable of being exercised towards any of a plurality of options that are in view for the agent. We may call the causing of this intentional state a ‘decision’ and suppose that in the usual case it is a triggering event, initiating the chain of events constituting a wider observable action.

Following agent causalists of tradition (such as Thomas Reid), I conceive agent-causal power as *inherently* goal-directed. It is the power of an agent to cause an intention in order to satisfy some desire or to achieve some aim. How should we understand this goal-directedness?

One way is a variation on a familiar view in action theory, the causal theory of action. The rough idea behind developed versions of the causal theory is that the agent's having a potential reason R actually motivated action A (and so contributes to its explanation) just in case R is a salient element in the set of causes that 'nondeviantly' produce A. (R might be, e.g., a belief–desire pair or a prior intention.) In line with this, one might be tempted to say that R actually motivated *an agent's causing of intention i* just in case R is among the set of causes that nondeviantly produce it.

However, it is not clear that anything *could* (in strict truth) produce a causally-complex event of the form *an agent's causing of intention i*. On the causal powers theory, causation consists in the manifestation of a single disposition (limiting case) or the mutual manifestation of a plurality of properties that are, in C. B. Martin's term, 'reciprocal dispositional partners' (typical case). We sometimes speak of an external event's 'triggering' a disposition to act (as when the lighting of the fuse is said to trigger the dynamite's disposition to explode). And this way of speaking might tempt one to assert, in more explicit terms, that the lighting of the fuse directly caused the causally-complex event, *the dynamite's emitting a large quantity of hot gas caused the rapid dislocation of matter in its immediate vicinity*. But neither of these statements can be taken at face value when it comes to the metaphysics of causation. The unstable chemical properties of the dynamite's active

substance (nitroglycerin) would be involved in any number of effects relative to an appropriate wider circumstance. However, in each case, a variety of conditions C are exercising a *joint* disposition towards the particular effect, E. And nothing directly *produces their producing* of the effect, as this could only mean that the conditions C did not, after all, include all the factors involved in producing E. Granted, there is a perfectly good sense in which a prior event that produces one or more of C's elements may be said to *indirectly* cause C's causing E, in virtue of causing part of C itself to obtain. (So, the lighting of the fuse did lead more or less directly to the rapid burning process of the chemical substance, which event caused the production of the hot gas, which in turn caused the surrounding matter to be rapidly pressed outward.) But this indirect causing of a causal chain by virtue of causing the chain's first element cannot apply to instances of agent causation, for the simple reason that here the first element within the causal chain is not an event or condition, but a substance. It is not the event of the agent's existing at t that causes the coming to be of a state of intention – something that *could* have a cause – but the agent himself. The notion of *causing a substance*, qua substance, has no clear sense.<sup>10</sup>

---

<sup>10</sup> See O'Connor 2000: 52–5, for further discussion.

Thus, we cannot coherently suppose that the obtaining of a reason in the agent may be said to be among the factors that causally *produce* the agent's causing an intention. But perhaps we can sensibly stop a little short of this supposition by supposing instead that the obtaining of the reason appropriately affects (in the typical case, by increasing) an *objective propensity of the agent* to cause the intention. On this latter suggestion, while nothing produces an instance of agent causation, the possible occurrence of this event has a continuously evolving, objective likelihood. Expressed differently, agent causal power is a structured propensity towards a class of effects (the formings of executive intentions), such that at any given time, for each causally possible, specific agent-causal event-type, there is a definite objective probability of its occurrence within the range (0, 1), and this probability varies continuously as the agent is impacted by internal and external influences. To emphasize: events that alter this propensity do *not* thereby tend toward the production of the agent's causing the coming to be of an intention (in the sense that they potentially contribute to the *causing* of this latter event). Even where the event promoted occurs, the effect of the influencing events is exhausted by their alteration of the relative likelihood of an outcome, which they accomplish by affecting the propensities of the agent-causal capacity itself. Where reasons confer probabilities in this manner, I will say that the reasons *causally structure* the agent-causal capacity.

It will perhaps be helpful to clarify what I have and have not just said. I am at this point taking it as provisionally given that we have a decent grasp on the very idea of agent causal power. I argued in the penultimate paragraph that we cannot coherently suppose reasons to constrain agent-causal power in one familiar way, that of tending to produce agent-causal events. I then tried to explicate another (albeit less familiar) way, that of structuring a propensity of an agent to produce an event *without thereby tending toward the production of* the agent's producing said event. However, I have in no way suggested that one could not coherently jettison the idea of agent causation in favor of an event-causal theory of action on which the having of reasons does indeed tend directly towards the production of one's executive intentions to act. This is the guiding idea of causal theories of action and, unlike some agent causationists such as Richard Taylor, I do not maintain that it is impossible to give a satisfactory theory of action along these lines. I would only insist that such a theory cannot capture the more ambitious notion of *freedom of action*. Here, I maintain, purely event-causal theories (whether deterministic or not) will inevitably fail. Thus, I am committed to supposing that there is more than one broad sort of way that the having of reasons might influence an intentional action. As with other propensities, the effect of events constituted by the having of reasons to act depends on surrounding circumstances. The agent-causal account I am advancing suggests that the presence of agent-causal power is one very

important determinant on such effects. In the presence of such a power, the causal contribution of the having of reasons is exhausted by the alteration of the probability of a corresponding agent-causal event.

With this idea in hand, we can specify *one* way in which agent-causal events are inherently purposive. Necessarily, when an agent causes an intention *i* to occur at time  $t_1$ , he does so in the presence of a motivational state whose onset began at a time  $t_0 < t_1$  and which had an appreciable influence on the probability between  $t_0$  and  $t_1$  of his causing of *i*. Let us say that when a reason *R* satisfies this description, the agent freely acted *on* *R*.<sup>11</sup>

The controverted metaphysics of agent causation aside, this sufficient condition for acting on a reason is quite minimal. Some would contend that it is objectionably weak because it allows that I may act on a reason of which I am entirely unconscious. In such a case, it will be claimed, the reason is exerting a brutally causal influence, and not a rational influence.<sup>12</sup> In my view, this objection mistakenly seeks to assimilate all cases of reasons-guided activity into a single

---

<sup>11</sup> I first developed the distinction in the text between *acting on* and *acting for* a reason in O'Connor 2005.

<sup>12</sup> Nikolaj Nottelmann raised this objection to me in discussion.

framework. I agree that *some* free actions manifest a heightened degree of conscious control and I try to capture what this might consist in immediately below. But we need to recognize that not even all free actions are created equal – freedom itself comes in degrees. If a reason inclines me to undertake an action but its content is unknown to me (if, say, I am aware merely that I have an inclination to undertake the action), the latter fact diminished my freedom, since I am thereby unable to subject my motivation to rational scrutiny. Nonetheless, if it remains open to me to undertake the action or not, I exhibit the goal-driven self-determination that is the core element of freedom of the will.<sup>13</sup>

Agents who act freely to any degree, then, directly produce the intentions that initiate and guide their actions, acting on an inclination that is the causal product of certain reasons they acquired (and subsequently retained) at some point prior to this causal activity. But sometimes, it would appear, there is more to be said about the way that reasons motivate freely undertaken actions. Often enough, not only am I conscious of certain reasons that favor the course of action I am choosing, I expressly choose the action for the purpose of achieving the goal to which those

---

<sup>13</sup> For further discussion of the idea of degrees of freedom, see O'Connor 2005.

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

It is commonly objected to nondeterministic accounts of human freedom that, despite what I've just said, undetermined actions cannot be explained by the agent's reasons since those reasons cannot account for why the agent performed

action A rather than B, one of the alternatives that were also causally possible in the circumstances. This objection fails to appreciate that explanation need not always be contrastive. If there are truly indeterministic quantum mechanical systems capable of generating any of a plurality of outcomes, whatever results is not absolutely inexplicable. A perfectly good explanation may be given by citing the system and its relevant capacities that in fact produced the outcome, even if there is no explanation at all of why that outcome occurred rather than any of the others that might have. Similarly, if an agent is capable of causing any of a range of intentions that would result in different corresponding actions, the reason(s) that inclined the agent to do what he in fact does serve to explain it even though there may be no explanation of why he did that rather than any of the alternatives.

## 2 Arguments for the Impossibility of Agent Causal Power

The possibility of agent causation has been widely doubted, especially in recent philosophy. Sometimes, the claim is that it is absolutely, or metaphysically, impossible insofar as it posits conditions that contradict certain necessary truths about certain ontological categories, such as that of event or substance. Other times, the form of impossibility is epistemic: agent causation, it is held, is incompatible with what we have excellent reason to believe are basic truths concerning the physics of our world; or with the relationship between the physics and any high-level processes to which basic physics gives rise; or with a general sort

Toby

**Comment:** I think the inserted clause here does not do enough to make it clear what the "minimal conditions" are that are of interest. I think a short paragraph could be inserted in the vicinity, briefly describing the sort of possibility that is of interest. I take it that you are interested in the epistemic possibility, given what we know about the actual world, \*and\* given certain constraints of metaphysical plausibility. That way you are ruling out weird worlds in which physics is totally different, and you are also ruling out worlds which are metaphysically extravagant, while having the same physics as the actual world. (?)

of ontological economy that obtains in our world (no superfluous explanatory features). I will now consider four reasons that some have put forward as grounds for doubting the coherence of agent causation, either absolutely, or with respect to some such general feature of our world.

*From the timing of actions*

C. D. Broad's oft-cited objection to the possibility of agent causation runs thus:

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

(1952: 215)

Broad's objection, or something like it, would have considerable force against an agent-causal view that maintained that nothing about the agent at the time of his action was explanatorily relevant to its performance. Such an "action" would indeed seem freakish, or inexplicable in any significant way. But no agent causationist imagines such a scenario. On the version of the view advanced here, the agent's capacity to cause action-triggering events is causally structured by the

agent's internal state, involving the having of reasons and other factors, before and up to the time of the action. These events within the agent suffice to explanatorily ground the agent's causing the event to happen "at a certain date" without collapsing the view into one on which those events themselves produce the action.

Randy Clarke, an erstwhile defender of an agent-causal account of freedom, has recently claimed that a modified version of Broad's objection has some force.<sup>14</sup>

Events, but not substances, are 'directly' in time in that their times are constituents of the events. By contrast, he maintains, "a substance is in time only in that events involving it...are directly in time." (This is supposed to be directly parallel to a reverse contention with respect to space, on which substances occupy space directly whereas events in their careers occupy a location only via its constituent object.) From this, he suggests, one can argue that the fact that effects are caused to occur at times "can be so only if their causes likewise occur at times –

---

<sup>14</sup> 2003: 201–2. Clarke does not claim that this argument is individually decisive. Instead, he presents several considerations, including some that I present below, that he believes to tell against the possibility of agent causation and that cumulatively make the impossibility of agent causation more likely than not. I will argue to the contrary that none of the main considerations he adduces has significant force.

only, that is, if their causes are directly in time in the way in which events are but substances are not” (2003: 201).

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

*From the uniformity of causal power*

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

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

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

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

We may read this objection as making the claim that the ontological category of *property* has an abstract functional essence that includes the tendency in the presence of other properties towards the direct, joint production of certain effects.

Is there reason to think that this is so? One reason to doubt it stems from the variety of property theories philosophers have advanced: transcendent vs. immanent universals vs. tropes; pure powers vs. Humean non-dispositional qualities vs. a ‘dual-aspect’ combination of the two. In the face of serious commitment to and elaboration of such diverse positions, one might think that the range of possibility is correspondingly broad. However, this sort of reason for rejecting the uniform dispositional essence thesis is not compelling. Nearly all philosophers holding one or another of the competing theories of properties just noted believes the truth of their favored theory to be necessary, with the other alleged entities judged to be impossibilities. And that seems the right thing to say, whatever one’s view. In particular, if one rejects Humean qualities in favor of either the pure powers or dual aspect theory, then one should hold that, whatever the variation across possibility space when it comes to the specific kinds of properties there are, they all conform to the general features of one’s property theory.<sup>15</sup>

---

<sup>15</sup> A complication here is that one version of the causal powers theory

Thus, one who maintains the uniformity thesis may allow that figuring out the right conception of properties is difficult while contending that, once we have embraced a general approach (here presumed to include irreducible dispositionality), we should presume an absolute unity of nature at a suitable level of abstraction. But at *what* level of abstraction should the thesis be applied? Consider that, in the advent of statistical laws in fundamental physics, many metaphysicians are now comfortable with the notion that there are nondeterministic dispositions varying in strength along a continuum, with deterministic potentialities merely being a limiting case. Consider further that, while properties typically work in tandem towards effects, a natural way of interpreting the phenomenon of radioactive particle decay is as an entirely self-contained process whose timing is radically undetermined by any sort of stimulus event. Finally, some adhere to the truth of (and still others to the possibility of) a view that all or many conscious mental properties are intrinsically intentional while this is true of no physical properties. None of these claims concern free will,

---

maintains that there are two fundamental kinds of properties and relations: those whose nature is dispositional and certain 'framework' relations such as spatio-temporal relations. (See Ellis 2001.) Suppose this is correct. Even so, one may argue for the abstract functional uniformity of all properties falling on the powers side of this division.

and yet all posit a kind of variability in the nature of dispositional properties that warrants classifying them into different basic types. Given these examples, it is hard to see why there may not be a further partition of types of the sort envisioned by the agent causationist. Doubtless there is a unity across these divisions at *some* level of abstraction. But assuming the agent causationist's position is otherwise motivated, he may reasonably contend that it must be sufficiently abstract as to encompass the division his theory requires. Indeed, why may not the unity of *basic* dispositional properties simply consist in their making a net addition to the pool of causal powers?

*From the connection between causation and probability-raising*

A third consideration for doubting the possibility of agent causation (also given by Clarke 2003) takes as its point of departure that causation is somehow bound up with probability raising. As Clarke notes, it is difficult if not impossible to state a very precise thesis here. In most cases of indeterministic event causation, the obtaining of a causally relevant feature to some degree *raised* the probability of the effect in question – but not, or at least not obviously, in all cases. Given the fact that the obtaining of potential causes may screen off other factors that would otherwise potentially influence an outcome, we can readily imagine cases where

something that actually contributed to an effect nonetheless rendered it either no more or even *less* likely than it would have been had that factor not obtained.<sup>16</sup>

Still, he contends, there is “considerable plausibility” to the claim that a cause must be the sort of entity that can *antecedently affect the probability* of their effects. But an agent, as such, is not this sort of entity – clearly, only an event or enduring state could be. So the thesis of agent causation runs contrary to a plausible constraint on indeterministic causation by positing agents who are indeterministic causes of certain effects while necessarily not having a direct influence on the prior probability of those effects (203).

In reply, I suggest that Clarke’s proposed constraint on admissible causes is at best a rough, first pass at capturing a conceptual connection between causal factors and probability transmission. And to the extent that there is a plausible intuition underlying it, my structured tendency account of agent causation conforms to it. On this account, agent causation is not something wholly disassociated from the evolving chain of probabilistic causes constituting the world’s history. Agent causes act on the basis of prior factors that confer a positive objective probability on their

---

<sup>16</sup> Clarke cites (203) putative illustrations in Dowe 2000: 33–40, Salmon 1984: 192–202, and Ehring 1997: 36. For a deft and thorough development of a theory of indeterministic event causal powers, see Hiddleston 2005a.

Toby 1/15/08 2:45 PM  
Deleted: )

occurrence. This should suffice to conform to any independently plausible, refined thesis that is intended to capture the vague intuition that probability transmission is a fundamental feature of causation. We may grant that, if one *already* believes on independent grounds that agent causation is impossible, then Clarke's more specific claim is a reasonable way to express the intuition. But in the present context, it seems to me a gratuitous strengthening of a general intuition in a way that arbitrarily precludes the possibility of agent causation.

*From the superfluity/unknowability of agent causing that is probabilistically constrained*

By constructing a picture on which agent causal power is causally structured by reasons and other factors, I have tried to integrate the view of human freedom it anchors into a view of the wider world as an evolving network of interacting powers. Critics of traditional, less constrained versions of the view have understandably complained that it depicts a godlike transcendence of natural forces. The present project suggests that the objectionable aspect is inessential, since it is possible to give freedom a human face.<sup>17</sup>

---

<sup>17</sup> Another recent effort along these lines is O'Connor 2005. Randolph Clarke 2003 is similarly an attempt to give a recognizably human version of the

To other critics, however, that merely opens up other vulnerabilities. Eric Hiddleston charges that the view of agent causation as probabilistically structured renders it superfluous for explanatory purposes (2005b: 552–3). Any event that one might explain as caused by an agent whose power was probabilistically structured by reasons R1 and R2 might equally well and more economically be explained by the direct causal efficacy of R1 and R2, acting nondeterministically. Furthermore, the agent causal theorist practically invites such an alternative explanation when he allows that in some cases, reasons do in fact bring about actions (though these would not be directly *free* actions).

An initial reply simply notes that the agent causationist will expect that cases where reasons directly bring about actions differ in detectable ways from those in which agent causal power is at work. Viewed from another direction, there is no reason to suppose that an agent whose agent-causal power was rendered inoperative with all else being left untouched would simply carry on exactly as before.

But Hiddleston, I believe, would wish to persist as follows: why would we have reason to posit agent causal power in the first place instead of a causal theory of

---

agent causal account of freedom. John Churchill and I give reasons for rejecting Clarke's approach in our 2004.

action on which reasons are productive? It seems that nothing in the observable pattern of events could, even in principle, require us to ascribe certain events to agents as causes.<sup>18</sup>

In reply, I first question the premise that there are no distinctive features of agent-causal processes as against possible cases in which reasons nondeterministically produce actions. First, just given the way highly deliberate and unpressured choices seem to us to unfold, it seems doubtful that our reasons fix a probability for the precise timing of the action which they promote. At most, there is an interval of time in which a possible choice is likely to occur, with the likelihood being subject to fluctuation as the agent deliberates. It seems, on some occasions, that I am capable of putting off decisions or continuing to search for reasons pro and con various courses of action, and that there are not fixed probabilities of my ceasing to dither at particular moments. (This is consistent with assuming that there is a high conditional probability for the agent's causing intention  $i$  at  $t_1$  on  $i$ 's occurring at  $t_1$ .) One might construct a model analogous to the way physicists model particle decay, on which a sample of a radioactive substance of a given mass has an

---

<sup>18</sup> Clarke (2003: 206) endorses the undetectability claim while denying that it gives reason to reject the possibility of agent causation.

associated probability distribution that measures the likelihood for each moment over an interval of its having lost half its mass by that time. But any such model generally applied to the case of freely making a choice would seem contrived.

Second, the agent causationist can plausibly suppose that while there is a probability of an agent's causing an intention to A, there is no particular probability of that intention's having a further goal-directed content in the sense I spoke of as acting *for*, and not just *on*, a reason. That is, there may be a probability of 0.6 that an agent will, within a specified interval of time, cause an intention to A, but no probabilities to its having the precise content of simply A, *A-for-the-sake-of-G1*, or *A-for-the-sake-of-G2*.

I put forward these two suggestions with some diffidence, as they are grounded in vague intuitions concerning how actions seem to us to unfold. A more firmly-grounded response to the 'no evidence' objection is that, even if there weren't these differences from possible cases of indeterministic causation by reasons, we have a general explanatory reason to accept that agents are sometimes causes. The thesis that agents are sometimes causes, in my view, best reflects the phenomenology of many of our actions and it is required by our (related) native belief that most human adults are free and morally responsible agents.

In general respects, the explanatory appeal to agent causation is similar to the appeal to causal realism. The skeptical Humean insists that all we need are the observable patterns among events, (allegedly) conceived nondispositionally. According to the realist, we must go deeper, explaining the patterns themselves in terms of the structured propensities of the world's basic features. Humean complaints that primitive causation is an 'occult' metaphysical relation are just so much throwing dust in the air and complaining that one cannot see. We have an intuitive grasp – albeit one that is highly general, needing elucidation – on the idea of causal power and it is fundamental to our naïve conception of the world around us. Now, Hiddleston himself is no Humean. As it happens, however, he endorses Nancy Cartwright's (1989) idiosyncratic claim to provide a purely *empirical*, neutral case for the existence of event-causal dispositions, a case that turns on the nature of successful scientific practice in isolating causal influences. (And so he thinks there is a principled asymmetry concerning the possibility of evidence for event and agent causal dispositions.) Here, I can only join hands with many others in confessing not to understand exactly how Cartwright's argument is supposed to go – how it is that the convinced Humean cannot give an alternate interpretation (however perverse to my realist eyes on non-empirical grounds) of what is observed and the empirical inferences and theorizing based upon it.

Many philosophers (including Hiddleston) dispute the claim that our experience of acting in any way supports the thesis that human agents are sometimes (literally) causes. Randy Clarke writes:

I do not find it a credible claim that ordinary human agents have any experience, or any belief arising directly from experience, the content of which is correctly characterized in terms of agent causation... Ordinary human agents, it seems plain, lack the concept of substance causation. A representation of free action as substance-caused (or as consisting on the substance causation of some event internal to the action) is a sophisticated philosophical construction... (2003: 206–7)

Now, something in the neighborhood of Clarke's remarks is surely correct. It takes philosophical reflection to attain a clear grasp on the idea of agents *qua* substance as causes, as distinct from either events internal to the agent as causes or the bogus Humean surrogate of patterns of (actual or counterfactual) regularity among events internal to the agent. But, for all that, it may be (and I contend, *is*) the case that (a) the content of the experience-in-acting of ordinary human agents involves a fairly inchoate sense of themselves as bringing about their actions and that (b) the reflective account that best captures this inchoate content is the agent-causal account. I observe that this position on the ordinary *experience* of agency is

supported by Daniel Wegner, a prominent cognitive psychologist who goes on to argue that our experience of agency is deeply illusory.<sup>19</sup>

*From the insufficiency (for freedom) of agent causing that is probabilistically constrained*

Recall that an apparent difficulty facing an alternative, causal indeterminist account of human freedom – an account that eschews agent causation in favor of a nondeterministic causation of choices by one’s reasons – is that it appears to be a matter of luck which of the undetermined possibilities is realized in a particular case. Given the presence of desires and intentions of varying strength, making certain outcomes more likely than others, the agent possesses no further power to determine which outcome in fact is brought about. The determination is a product of the propensities of the agent’s states, and the agent doesn’t seem to directly control which propensity will ‘fire.’ If we imagine two identical agents in identical circumstances, with one agent nondeterministically choosing alternative A and the other choosing B, it seems a matter of luck from the standpoint of the agents themselves which alternative occurs in which person.

---

<sup>19</sup> See Wegner 2002, chapter 1. I criticize Wegner’s case for the illusory character of what he terms “conscious willing” in my 2005: 220–6.

Supposing there is a power of agent causation has the virtue that it seems to avoid this ‘problem of luck’ facing other indeterministic accounts.<sup>20</sup> Agent causation is precisely the power to directly determine which of several causal possibilities is realized on a given occasion. However, Derk Pereboom has recently argued that this is so only if agent causation does not conform to pre-given indeterministic tendencies.<sup>21</sup> He writes:

...to answer the luck objection, the causal power exercised by the agent must be of a different sort from that of the events that shape the agent-causal power, and on the occasion of a free decision, the exercise of these causal powers must be token-distinct from the exercise of the causal

---

<sup>20</sup> In addition to causal indeterminist and agent causal theories of freedom, there is noncausal indeterminism, on which control is an intrinsic, noncausal feature of free choices or actions. For versions of this view, see Goetz 1988, Ginet 1990, McCann 1998, and Pink 2004.

<sup>21</sup> Others have recently argued that agent causation does not necessarily avoid the problem of luck (or, as it sometimes put, the “problem of control”). See Haji 2004, Widerker 2005, and Mele 2006. I will not here address their ways of pressing the issue, though I find their arguments even less compelling than the one by Pereboom that I discuss in the text. Pereboom (2005: 243–4) also rejects the arguments of Mele and Haji.

powers of the events. Given this requirement, we would expect the decisions of the agent-cause to diverge, in the long run, from the frequency of choices that would be extremely likely on the basis of the events alone. If we nevertheless found conformity, we would have very good reason to believe that the agent-causal power was not of a different sort from the causal powers of the events after all, and that on the occasion of particular decisions, the exercise of these causal powers was not token-distinct. Or else, this conformity would be a wild coincidence...(2005: 246)

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

The agent causationist readily endorses the first of these conditionals, on a straightforward reading of “different sort of power” and “token-distinct exercise.” After all, the view posits a fundamental, irreducible power of agents to form intentions. But the second conditional directly rejects the viability of any account on which agent causal power is probabilistically structured by reasons. Why does Pereboom assert it? His thought seems to be that if the event of one’s having certain reasons along with other prior events ensure that one’s choices will fit a certain pattern – more accurately, make the pattern-fitting likely, given a sufficiently large number of cases – , then one’s supposed agent-causal power in choosing is at best a shadowy accompaniment to the event-causal power. In truth, it is no power at all, as it adds nothing to the mix of factors already in play. With no authority to act on its own, its presence makes no discernible difference to what occurs in the aggregate. If it would be a matter of luck, beyond my direct control, which of my indeterministic propensities happens to be realized on any given occasion, were the causal indeterminist account correct, then adding the ability to ‘directly determine’ the outcome wouldn’t help if I am ineluctably constrained by those very propensities.

It is easy to feel the pull of this thought, but it should be resisted. First, we insist upon the importance of the distinction between (the persisting state or event of one’s having) reasons *structuring* one’s agent-causal power in the sense of

conferring objective tendencies towards particular actions and reasons *activating* that power by producing one's causing a specific intention. On the view I have described, nothing other than the agent himself activates the agent causal power in this way. To say that I have an objective probability of 0.8 to cause the intention to join my students at the local pub ensures nothing about what I will in fact do. I can resist this rather strong inclination just as well as act upon it. The probability simply measures relative likelihood and serves to predict a distribution of outcomes were I to be similarly inclined in similar circumstances many times over (which of course I never am in actual practice). The reason that the alternative, causal indeterminist view is subject to the luck objection is *not* that it posits objective probabilities to possible outcomes but that it fails to posit a kind of single-case form of control by means of which the agent can determine what happens in each case. After all, were the causal indeterminist picture modified so that agents choices are caused but not determined by appropriate internal states whose propensities, while nondeterministic, lacked definite measure, the problem of luck or control would remain. Again, that problem concerns not prior influence but the ability to settle what occurs on the occasion of a causally undetermined outcome. The agent causationist's solution is to posit a basic capacity of just that sort, while allowing that the capacity is not situated within an indifferent agent, but one with evolving preferences and beliefs. Surely having preferences does not undermine control!

Toby 1/15/08 2:45 PM

Deleted: a

Causation is a primitive, yet fairly simple concept. Reflective theorizing, both philosophical and scientific, has yielded a variety of forms in which it can be coherently conceived. Absent convincing reason to think one of these imagined forms is **defective** in some unobvious way, we ought to deem it possible. It has been the aim of the present essay to specify a coherent way we might think about the idea of agent-causal power, an idea that has guided some who theorize about freedom of the will and that is even deemed intuitively attractive by some who resist it. It may be metaphysically impossible that there be such a thing, but I have yet to encounter a convincing argument for it.

Toby 1/15/08 2:45 PM

**Deleted:** defective