CHAPTER 3

Free Will and Metaphysics

TIMOTHY O'CONNOR

For almost fifty years, Robert Kane has been a man on a mission. While he has addressed a range of questions in metaphysics and value theory throughout his distinguished career, he has sought with more tenacity than any other living philosopher to clarify the metaphysical underpinnings of human freedom and moral responsibility. Kane began thinking about the topic in the early 1960s, and I think it is worth noting that era's climate of opinion regarding not just free will (something he frequently notes) but also metaphysics and the philosophy of mind more generally. While logical positivism was on the wane, empiricist suspicion of traditional metaphysical categories and arguments endured. Since that time, metaphysics has been re-born and is now flourishing. Most pertinent to the present essay, the metaphysics of causation and the ontology of mental states have returned as hotly debated issues.

The central theme of the present essay is that an adequate account of free will must squarely engage these more fundamental metaphysical issues—they cannot be "bracketed off" in the way that even some contemporary theorists of free will tend to suppose. Doing so involves considering both empirical and philosophical issues raised by our fundamental physical theory, quantum mechanics, and how the processes it describes connect to the macro-level phenomena in the brain and mind. Interesting developments have occurred on this front, and Kane has tried to appropriate some of them in his thinking on free will. Some theorists have offered schematic models of how indeterministic quantum effects might be amplified in brain processes underlying human decision-making. More ambitiously, I think,

has been the general rethinking of reductionist metaphysics on which all macro-level processes are either identical to or wholly constituted by micro-level processes. Although scientific theorists' ideas can be difficult to interpret in terms of philosophical categories, there has clearly been some form of anti-reductionism at work in much thought arising from complex systems theory, applied to physics, biology, and elsewhere. In philosophy of mind, there has been a reappreciation of the challenge to reductionism posed by consciousness and (more recently) the intentionality of mental states.

I agree with (and here assume) Kane's basic orientation to the problem of free will, on which metaphysical freedom consists in being the ultimate, reasons-guided causal source of an intention to act in the face of alternatives possibilities for action. Over the years, he has developed and refined a complex analysis of what the exercise of such ultimate causality consists in, an analysis that is intended to contrast with "agent-causal" accounts that take it as a conceptual and metaphysical primitive. I will argue, however, that when we draw out two metaphysical assumptions to which Kane's account is plausibly committed, we can easily be led to an account that is just a particular version of something like the primitivist agent-causal theory itself. Put differently, when set within a plausible general metaphysical framework, Kane's theory and the agent-causal theory are much closer than has so far been recognized.

1. KANEAN LIBERTARIANISM

Central to Kane's analysis of free will is the notion of "will-setting" or "self-forming" actions (SFAs). As many recent action theorists (and for that matter, cognitive psychologists) emphasize, much of our behavior is automatic, unfolding in accordance with entrenched action plans that are triggered, in some cases entirely unconsciously, by the appropriate stimuli, without any intervening choice or active intention-formation. For example, you are in your car at a stoplight thinking about the puzzle of free will. The light turns green and, without thinking about it, your foot moves from the brake to the gas pedal. However, the interesting cases are ones in which we do consider what to do and feel some pull in more than one direction. These might be overtly moral choices, where there is conflict between "duty" and "desire," or choices between short- and long-term self-interest, or simply choices among a range of equally permissible and prudent actions that are all worthwhile from the agent's point of view but not all of which can be undertaken. It is these cases, Kane believes, where it is most plausible to suppose that agents directly exercise their freedom of will. They are

cases when the agent's own will is divided between incompatible courses of action, in the sense that she has a plurality of "volitional streams"—complexes of beliefs, desires, and intentions—that are aimed at different ends. The agent is *trying* to accomplish each of two or more incompatible goals, and hence is divided. He posits that there is an objective, nonnegligible chance that each of these efforts succeeds. Whatever the outcome of this internal struggle, it will have been the agent's own effort that has brought it about, a choice that is both motivated and intended. Finally, as Aristotle taught, over time, individual choice outcomes affect the relative weight of subsequent volitional streams, and a character is formed that is partly a result of the agent's previous choices (Kane 2011a, pp. 386–390).

2. TWO METAPHYSICAL COMMITMENTS OF KANE'S ACCOUNT

2.1 Ontological Irreducibility of Mental States

Kane rejects the thesis that human persons are immaterial minds. I concur, but this negative thesis is consistent with a wide range of views concerning the nature of mental *states*. According to standard forms of both reductive and (putatively) "nonreductive" physicalism, *token* mental events supervene upon (and on most versions, are identical to) structured physical events in the brain. They differ in that nonreductive physicalism denies, while reductive physicalism affirms, that they are type identical. Jaegwon Kim has argued in numerous writings (see especially Kim 1998) that nonreductive physicalism is an untenable position, in that it is unable to secure the causal efficacy of mental events. While there are problems with Kim's presentations of the argument, I believe that there is a sound version of it (O'Connor and Churchill 2010). And reductive physicalism is obviously an unsatisfactory position for one who affirms a libertarian position regarding free will. A more specific problem than Kim's challenge for both types of physicalist view of mental states is that they appear to be subject to a Consequence-style argument for their incompatibility with free will (Cover and Hawthorne 1996, pp. 58–60): for an arbitrary action A, let "P" be a proposition describing each of the constituent microphysical states and relations thereof that (according to physicalism) constitute my deciding to A at time t, and let "Q" be the proposition that I decide to A at time t. Plausibly, P and I have no choice whether P; and necessarily, if P then Q (by supervenience); hence Q, and I have no choice whether Q.

The upshot, I suggest, is that one who affirms Kane's position on the nature and reality of free will must reject all varieties of physicalism. Given an antecedent rejection of mind-body (substance) dualism, this result points

in the direction of a metaphysical form of emergentism. Kane (2011a) appears to concur (p. 396), but we need to be careful here. The term "emergence" is used frequently in relation to complex systems of all kinds. It is important that we distinguish metaphysical emergence theses from those that are, in one sense or another, merely epistemic.

Conway's simple cellular automaton, the *Game of Life*, vividly illustrates that the existence of strikingly novel *patterns* of behavior in complex macroscopic systems is consistent with the behavior of those systems being wholly determined by completely general low-level transition rules at the fundamental level. These high-level patterns are "emergent" only in the sense that one cannot—at least in any straightforward way—*derive* the patterns from the properties and patterns appropriate to the fundamental level alone; it remains the case that the high-level rules do not in any way *modify or supplement* the basic dynamics that drive *Life*'s evolution. In short, the transparent simplicity of *Life* worlds make plain to an observer that epistemic irreducibility (in some sense) of high-level patterns is consistent with metaphysical reductionism. It is true but misleading to say of *Life* worlds that cells caught up in stable macroscopic structures that follow different dynamical patterns are "constrained by" those high-level dynamics. Ultimately, what is happening is that individual cells *constrain themselves*, by "causing" there to be (in certain regions) such macroscopic structures and determining, moment by moment, what the precise state of those structures will be. That one can focus outward from those details and see large-scale patterns requiring a different form of description does not change the fundamental point that there is an asymmetrical dependency of macro-level patterns (where they occur here and there) upon the completely general micro-level patterns.

In contrast to the epistemic emergence reflected in the *Life* game, what is needed for freedom of the will in stable systems such as ourselves who are wholly physically composed is a *metaphysical* form of emergence: our mental states and capacities must be ontologically *basic*, rather than token identical to complex physical states, making a nonredundant causal difference to the way we behave. On such an emergentist picture, certain of our conscious mental states (perceptual, cognitive, and conative) are ontologically basic states of enduring, though changing, biologically composed objects that causally contribute to our unfolding mental and physical behavior, and specifically to the states of intention or decision whereby we resolve deliberative uncertainty and embark on courses of action.

It is sometimes suggested that there being metaphysically emergent capacities would be "spooky," not amenable to empirical investigation. But this is simply not the case. While they are basic features of reality, emergent

capacities may nevertheless be fruitfully studied and eventually explained in detail in nonreductive fashion, by spelling out the basic inventory of emergent properties, detailing the precise conditions under which organized physical systems give rise to them, and isolating the precise behavioral impact their presence has on the system. Where we have reason to believe there are such metaphysically emergent capacities, it will be natural to suppose that they are *caused* to be by the object's fundamental parts, which have latent dispositions awaiting only the right configurational context for manifestation. If, in human beings, the capacity to form choices that emerges operates indeterministically, its existence and causal nature could be studied in more fundamental physical terms, even though its outputs cannot be explained in purely physical terms.

I said above that Kane appears to affirm emergentism regarding some or all conscious mental states, but the details of his suggestions concerning how things might go in the processes constituting freely willed choices make it unclear to me whether he has in mind *Life*-style epistemological emergence or a robust metaphysical emergence. On the one hand, he suggests that there might be mechanisms whereby local micro-indeterminacies in relevant aspects of the brain might get amplified, determining which of two large-scale competing neural networks that encode opposing motivational structures has its characteristic end realized in action (1996, pp. 130–142; 2011a, p. 387). This is perhaps most naturally understood as a special case of a *Life* scenario: micro-level processes naturally result in the formation of stable structures that "constrain" individual components, and the outcomes of these structures are determined non-linearly and in a way that is sensitive to small-scale and relatively localized indeterminacies. On the other hand, he speaks more generally of our exercising "macro-control of processes involving many neurons" (2011a, p. 395). While not wanting to discount the potential contributing role of processes described by the former suggestion, I contend that it is crucial to the viability of Kane's claim that it is the agent herself that is controlling this outcome, that the central determinants of choice be metaphysically basic, agent-level powers.

2.2 Causal Nonreductionism

A second plausible commitment of Kane's account of free will (and one that was implicit in remarks above) is a realist, nonreductionist view of causation. Consider the neo-Humean reductionist view as developed by David Lewis. (I discuss Lewis's picture only for the sake of concreteness;

the incongruity with Kane's account of free will that I allege generalizes to any reductionist account.) According to it, causal facts and the laws of nature are reducible to facts concerning the global spatiotemporal arrangement of fundamental natural properties which we (allegedly) may conceive in nondispositional terms. Roughly, the laws are the best system of generalizations over such natural facts, where what is best is determined by the optimal balance of simplicity and explanatory "strength." Causation in turn consists of a restricted kind of counterfactual dependence of one event on another, where the counterfactuals are grounded in cross-world similarities.[1] Within this framework, intentional human agency is naturally understood in terms of the counterfactual dependence of behavior or behavior-guiding intentions on appropriate beliefs, desires, or intentions the agent had immediately before and as the behavior occurs.

This reductionist metaphysics of causation yields an implausible understanding of the metaphysics of agential control. By taking the fact of A's being a cause of B to be a reducible, massively *extrinsic* relation—grounded in what occurs elsewhere and elsewhen—we empty the fundamental idea that A "produces" or "brings about" B of any clear content. Since agency is a causal notion, the implausibility carries over: on a neo-Humean analysis, the sense in which my beliefs and desires here and now *bring about* my present action is at best very weak tea. *A fortiori*, extrinsic analyses, on which whether or not psychological factors cause behavior is largely determined (metaphysically) by what happens in the distant reaches of spacetime, provide a bizarre account of a *free* action's being, as we commonly say, "directly controlled by" the agent, such that it was "up to her" what she would do in the particular circumstances. Our notion of agential control, and especially the ultimacy condition on freedom of the will, manifestly indicates something that supervenes on the local circumstances in which we act. Freedom of the will cannot survive a reductionist construal of causation, of which it is a particular form.

The best alternative to a neo-Humean reductionist account of causation is a neo-Aristotelian causal powers account.[2] On this account, the "natural" properties of objects are causal powers—power to bring about particular results in particular circumstances constitute their fundamental, intrinsic

1. The locus classicus is Lewis (1986). (I note that Lewis allows for temporally remote causation by defining causal chains in terms of stepwise counterfactual dependencies, but it is unnecessary to fuss about such details here.)

2. I here ignore the higher-order laws account proposed in different forms by Tooley (1987) and Armstrong (1984). There are well-known, quite fundamental problems with these accounts.

nature.[3] Causation is the manifestation of such a power (or the collaborative manifestation of multiple powers in interacting objects). Indeterministic causation, no less than deterministic causation, is the manifestation of causal power, though here the power is associated with propensities that are "chancy" in the sense that the objective prior probability in a given circumstance that the power will be manifested as it in fact is is less than 1. These are propensities toward a plurality of possible effects, and thus propensities that are manifested in different ways on different occasions. Indeterministic causal powers are sufficient, relative to a context, for each of the possible outcomes in the sense that they are all that is needed, though not in the sense that they are a causally sufficient condition. Every indeterministic event is produced, though none is necessitated.

3. NEO-ARISTOTELIAN FREEDOM

In the previous paragraph, I have deliberately elided mention of the entity that is the *cause*—that which exercises the causal power. Within the broadly neo-Aristotelian framework, there are two different ways one might think about this issue that have significantly different implications for how we think about the metaphysics of freedom. According to the first, we should say that, in a given determinate situation type S, *the having of a power P by object O1 at time t* produces effect E in object O2. Or, perhaps more commonly, in situation type S, the having of power P1 by object O1 and the having of power P2 by object O2 *jointly* produce effect E.[4] That is to say, causes are *events*. The second analysis maintains, instead, that in situation type S, *the objects O1 and O2* jointly produce effect E, doing so *in virtue of* their having powers P1 and P2 at time *t*, respectively. They jointly exercise their respective powers P1 and P2 to contribute to bringing about E. That is to say, causes are *objects/substances*.

Something like the first, event-causal understanding of causation is implicit in Kane's discussion. That is unsurprising, since event-causal analyses of some form or other have been popular ever since Hume, and especially throughout the twentieth century. However, one may argue that the general identification of causes with events is a legacy of the Humean rejection of causal power and substance. Abandon these Humean deflationary projects and it becomes natural to understand causes as substances. On the

3. See, e.g., Shoemaker (1980), Heil (2003), Mumford (2004), Lowe (2008), Martin (2008), Bird (2010), and Jacobs (2011). Some authors say instead—misguidedly, I judge—that properties of necessity *confer* causal powers on their bearers.

4. The mutuality of causal interactions is much emphasized by Martin (2008).

neo-Aristotelian metaphysical framework, the world is fundamentally a world of things/substances, not events. Events are derivative from (constructed in part out of) objects. It is, in general, powerful particulars—objects—that exercise causal power, that do things in the world. To be sure, their acting in the ways that they do has an explanation: they reflect the causal powers that they have at the time, powers that are none other than (one or more) natural properties that they have, and also (typically) the presence of necessary manifestation conditions. (I say "typically" since the phenomena of radioactive particle decay seems to involve no manifestation conditions.) So, for example, two electrons, eddie and eleonore, mutually repel each other—that is, cause each other to accelerate along receding paths at a specific rate. They do so in virtue of their powers—that is, negative electric charge—and in the circumstance (necessary manifestation condition) of their being a certain distance apart.

This description of the metaphysics of causation is natural within the neo-Aristotelian framework. If we accept it, we have a significant benefit for the problem of free will: the agent causalist's "problem of the disappearing agent" worry concerning event causalist accounts of free will, such as Kane's, melt away. Since all causation is substance causation, then (provided we have a nonreductive view of agents and their powers, per our first assumption) unreduced "agent causation" comes for free—it is not a *fundamentally* distinct variety of causation.

If all this is right, then a Kanean "event causal" libertarian really is (or ought to be) a kind of agent causalist. I am, quite literally, the principal cause of my free choices—which choices involve significant macro-level indeterminism (of the right sort), dual rationality (*weakly* understood, requiring only that whichever choice I make, it would be motivated), and dual control (whichever choice I make, it be one that *I* bring about). Kane's distinctive notion of opposing "efforts of will" is also compatible with this general metaphysical framework, and may be argued on its own merits.