

Persons and Causes

The Metaphysics of Free Will

TIMOTHY O'CONNOR

OXFORD
UNIVERSITY PRESS

Contents

Introduction xi

1. *Freedom and Determinism 3*
 - 1.1 An Ancient Argument 3
 - 1.2 Some Modal Principles and the Argument for Incompatibilism 5
 - 1.3 The Fixity of the Past and of Natural Laws 15
 - 1.4 Freedom and Responsibility 18
2. *Freedom and Indeterminism: Some Unsatisfactory Proposals 23*
 - 2.1 The Trouble with Incompatibilism 23
 - 2.2 Simple Indeterminism: Carl Ginet on Choice and Control 24
 - 2.3 Causal Indeterminism: The General Strategy and a Problem Posed 27
3. *The Agent as Cause: Reid, Taylor, and Chisholm 43*
 - 3.1 The Agency Theory 43
 - 3.2 Thomas Reid 43
 - 3.3 Richard Taylor 49
 - 3.4 Roderick Chisholm 55
 - 3.5 Summary 60

Appendix: Chisholm's Later Writings on Agency 61
4. *The Metaphysics of Free Will 67*
 - 4.1 Overview 67
 - 4.2 Event Causation 68
 - 4.3 Agent Causation 71
 - 4.4 C. D. Broad's Objection to the Very Idea of Agent Causation 74
 - 4.5 Remarks on a Contemporary Alternative Account of Agent Causation 76
 - 4.6 Ersatz Agent Causation? 79
 - 4.7 Alternative Possibilities, Responsibility, and Agent Causation 81

5. *Reasons and Causes* 85
 - 5.1 Reasons Explanation and the Agency Theory 85
 - 5.2 Objections to the Account 86
 - 5.3 Reasons and Contrastive Explanation 91
 - 5.4 Reasons and Tendencies to Act: A Residual Problem and a Speculative Proposal 95
 - 5.5 When Is the Will Free? 101
 - 5.6 Conclusion 107

6. *Agency, Mind, and Reductionism* 108
 - 6.1 Introductory Polemics: 'The Emerging Scientific Picture of the World' 108
 - 6.2 Emergence 110
 - 6.3 Emergence and Consciousness 115
 - 6.4 The Emergence of Active Power 121
 - 6.5 An Epistemological Objection to Agent Causation 123
 - 6.6 Conclusion 125

- References* 127

- Index* 133

Introduction

The topic of this study is one of the oldest, most contentious, and most difficult topics in philosophy. That it should prove to be all of these things is itself very puzzling, at first glance. For the goal is merely to make explicit our everyday picture of ourselves as agents who adopt specific purposes in freely choosing how we shall act, choices that trigger and help sustain our actions. If virtually all of us think of ourselves as freely acting, purposive beings—when out living our lives, if not always when reflecting on the matter in our studies— why haven't philosophers of the past managed to bequeath to us a perspicuous and immediately recognizable articulation of that thought? Granted, whether any such philosophical account answers to the facts of the springs of ordinary human behavior is an open empirical question: the truly puzzling matter is that there should remain deep controversy over what empirical researchers should be *looking for* to answer the question.

Like other enduring philosophical conundrums, the problem of understanding the idea of free, purposive, responsible activity (free will) is difficult in part because it touches on other fundamental ideas: causation, explanation, and the nature of intentional states such as beliefs, desires, and purposes—just to start. Another source of difficulty is that we tend to assume that free agency admits of degrees. Philosophers not only disagree about the scope of free will in ordinary human beings (as opposed to God). They also dispute *how* genetics and environment influence a person's freedom of action. That is, once one fixes a basic concept of free will, it remains puzzling how free will so understood can be qualified.

The two features of the problem I've just mentioned pose difficulties for the individual philosopher in coming to his own view on the nature of free will (and so, a fortiori, for the community of philosophers to come to a shared view). A final obstacle directly to the goal of consensus stems from the fact that this project is rarely pursued in isolation. Philosophers want to do more than paint a commonly held picture of ourselves. We also want to put forth a vision of human beings and their place in the wider scheme of things. And for some (no doubt

most, to some degree), this will be a vision that competes with the conventional one. Clearly, our view of whether human beings act freely—and to what degree—is right at the center of our wider philosophical vision. Freedom of will is directly connected to the possibility or significance of moral responsibility, autonomy, the uniqueness of persons (involving creativity, originality, and their life histories in general), dignity, love, and friendship. In short, it is connected to everything that fundamentally matters to us in our relationships with one another.¹

There is at present a widespread trend in English-speaking philosophy toward ‘deflationary’ analyses of most traditional targets of philosophical inquiry. Although this project is all the rage, carried out under the banner of “Philosophical Naturalism,” it is itself rarely made the subject of explicit articulation—what precisely qualifies and why it should be undertaken in the first place. The general, if vague, impetus is to analyze philosophical notions in a way that makes them hospitable to a ‘naturalistic’ view of human beings that has apparently been handed down to us by ‘Science.’

My own approach to the subject of free will in this work is quite counter to the fashion for apologies on behalf of ‘Naturalism.’ I am driven in the first place to clearly understand the prephilosophical view of human agency, let the chips fall where they may concerning its compatibility with ‘Naturalism.’ But I am no more satisfied with mere conceptual analysis than are Naturalists. I, too, want to understand human nature as it is—its frailties, as well as its glories—and recognize that empirical work in relevant branches of psychology and biology will, in the end, have quite a lot to say about that. But we can assess the significance of the verdict of the relevant science—when it comes—only if we have already reflected on our values and their requirements. It is ‘wretched subterfuge’ to settle on a scaled-back notion of what we value about ourselves and then declare victory when—surprise!—it is patently clear on a little reflection that no empirical work is ever likely to undermine that conception.

So how is it that we prereflectively think of ourselves? Chapter 1 constitutes a first step toward an answer to this question. In it, I consider the issue generating the most basic divide among theorists of free will: roughly, whether or not free will is compatible with the thesis that human choice and consequent behavior is a causally determined outcome of antecedent factors in and around the agent. In contemporary philosophy, this perennial debate has centered around the validity of a certain ‘modal’ style of argument for incompatibilism that turns on principles concerning the logic of ‘unavoidability’ (as in ‘Brian’s kicking his sister was unavoidable for him at the time, given his unfortunate affliction with sibling aggression syndrome’). Although the incompatibilist argument is easily stated, a number of subtle issues connected to it have been insufficiently understood by previous discussants. My exploration of these matters suggest two conclusions: (1) some current formulations of the incompatibilist argument are clearly invalid, but (2) they are naturally repaired in a manner that restores intuitive assent. I then defend

1. The most comprehensive recent discussion by a philosopher of these conceptual interconnections is chapter 6 of Robert Kane’s excellent book *The Significance of Free Will* (New York: Oxford University Press, 1996).

the validity of my favored version of the argument against common compatibilist replies.

If, as I contend, free will requires that our choices not be causally determined, then a positive account of it must show how the causal ‘loose fit’ between antecedent factors (including the agent’s having reasons for various alternatives) and choice does not reduce to an injection of mere randomness in the springs of action. One needs to show how certain causally indeterministic mechanisms would confer (or contribute to) a kind of power or control over one’s own choices that would be lacking in any sort of deterministic scenario. I consider, in chapter 2, four such accounts that I believe to be wanting. I first consider the broad position of *simple indeterminism*, in which the agent’s control over, or determination of, his own choices is held to be entirely noncausal in character and is instead a consequence of intrinsic, noncausal features of the choice itself. I argue against this position in the course of examining the account of its most prominent proponent, Carl Ginet. I then turn to the thesis of *causal indeterminism*, in which the agent’s control over his own choices resides in the indeterministic (‘chancy’) causal efficacy of his reasons for so choosing. The idea here is that in every free choice, several options have a nonzero probability in the circumstance, and the actual outcome in every case will be caused by factors that prominently include the agent’s reasons for so choosing. A *prima facie* problem for this position is to explain how the agent directly controls the outcome in a given case. There are objective probabilities corresponding to each of the possibilities, but within those fixed parameters, which choice occurs on a given occasion seems, as far as the agent’s direct control goes, a matter of chance. I examine three versions of causal indeterminism that try to overcome this objection. Robert Nozick’s strategy is to characterize choice as ‘self-subsuming,’ in that one’s choice may itself conform to the very values reflected in the choice. Storrs McCall argues that the intentional explanation of choice is independent of any probabilistic causal explanation. Finally, Robert Kane focuses on the deliberative process that gives rise to (what he takes to be) paradigmatic cases of freedom of will—cases in which the agent struggles to prioritize conflicting values and desires. Kane suggests that the agent’s control over the outcome consists in (1) the agent’s close identification with each of the conflicting sets of motivations and (2) the activity of the ‘self-network,’ a stable network of values, preferences, and so on that constitutes the agent’s character at the time of acting. I argue that all of these strategies fall short, despite the considerable ingenuity each displays (and the genuine insights on particular issues that one can glean from their efforts).

Indeed, by the end of this chapter, it becomes clear that both simple indeterminism and causal indeterminism founder because they try to establish a kind of control distinctive of free and responsible agents in the absence of correspondingly distinctive ontological resources. This gap is precisely what the traditional (and, nowadays, notorious) notion of ‘personal’ or ‘agent’ causation is intended to fill. Near the end of his instructive and thoughtful book, *The Significance of Free Will*, Robert Kane skeptically allows that “maybe theories of agent-causation can be resuscitated. But the burden of proof must be on anyone who would do so” (p. 195). I accept this assessment, and in the remainder of this book, I try to discharge that burden.

I begin, in chapter 3, by examining in some detail the three most influential accounts of free will that make recourse to the notion of agent causation: the eighteenth-century Scottish philosopher Thomas Reid and the contemporary philosophers Richard Taylor and Roderick Chisholm. In the discussion of each, I argue a central thesis (a point of agreement in the case of Reid, but one of disagreement with both Taylor and Chisholm). With Reid, I emphasize that an instance of agent causation (“exertion of active power,” to use Reid’s term) is not prior to or logically independent of the intention that is the agent’s immediate effect. Against Taylor, I argue that an exertion of active power cannot itself be causally produced (although a wide range of factors are of course necessary for an agent to have such a causal capacity at the time of its exercise). Finally, contra Chisholm, I argue that recognizing that an agent’s exertion of active power is intrinsically a direct exercise of control over one’s own behavior suffices to dispel the worry that the agency theory must somehow tell a further story to explain how the agent controls this event.

These three contentions constitute the initial ‘fixed points’ from which my own articulation of the metaphysics of agent causation is developed in chapter 4. I there contend that we should think of causal notions in general in terms of the notion of ‘causal power’ or ‘causal capacity.’ Given this more general ontological framework, the difference between *event* (or broadly mechanistic) causation and *agent* (or personal) causation concerns the way in which causal capacities are exercised. With event causation, the capacity to generate a particular effect (in deterministic cases, that effect will be only one of a range of possible effects) is exercised ‘as a matter of course’: having in the right circumstances the cluster of properties that ground the capacity directly gives rise to one of the effects within its range. By contrast, having the properties that subserve an agent-causal capacity doesn’t produce an effect; rather, it enables the agent to determine an effect (within a circumscribed range). Whether, when, and how such a capacity will be exercised is freely determined by the agent. After responding to some objections to the coherence of this basic account, I critically examine Randolph Clarke’s alternative account of agent causation, as well as some recent ‘deflated’ (*ersatz*) accounts.

In chapter 5, I defend an account of how agent-causally generated activity may be explained in terms of reasons. Central to the account is the assumption that what an agent directly causes is an action-triggering state of intention. The content of that intention (which endures throughout the action and guides its completion) is that an action of a specific sort be performed for certain reasons the agent had at the time. It is this twofold *internal* relation of direct reference and of similar content that grounds the explanatory link. I then respond to several objections stemming from Donald Davidson’s influential critique of noncausal accounts of reasons explanations. Along the way, I defend the position that such explanations need not be *contrastive* in character—that is, whereas there may be a reasons-based explanation of the agent’s acting as he did, this does not entail that there will also be an explanation of his so acting rather than performing any of the alternatives that had been contemplated. Near the end of the chapter, I suggest a refinement of the basic account in which the having of reasons generates or raises a *carried tendency* to act in particular ways, which tendency probabilistically structures the

basic agent-causal capacity. It remains up to the agent, nonetheless, to determine which such tendency will be acted on. This refinement of the basic account allows both (1) a straightforward interpretation of the fact that we have relative tendencies to act, even when apparently acting freely, and (2) an account of strongly contrastive, as well as noncontrastive, explanations of actions.

Finally, in chapter 6, I discuss the assumption of many 'Naturalists' that this agent-causal account of free will is not consistent with 'the emerging scientific picture of the world.' I contend that there is little basis for this claim. We must sharply distinguish the plausible claim that macrophysical phenomena in general arise out of and are causally sustained by microphysical phenomena ('The Causal Unity of Nature Thesis') from the far from evident claim that all such higher level phenomena are constituted by more fundamental, lower level phenomena ('The Micro-Macro Constitution Thesis'). Granted the former thesis, why accept the latter? It does not follow from the former, as is shown by the possibility of some higher level features being emergent in a robust sense. And when the issues are properly sorted out, I argue, it is plain that the Constitution Thesis is not empirically established. To be sure, difficult issues concerning the underlying basis for "active power" would have to be sorted out before a decent conception of it as emergent is to be had. But as things now stand, such a construal isn't precluded by present knowledge. So although agent causation may be widely disdained by Naturalists, it is not at odds with naturalism.