Introduction

Human persons are natural organisms. They are also morally responsible agents. They have, within limits, power to change themselves and their surroundings, and they may sometimes rightly be held responsible for the way they exercise this power. Some parts of nature (human beings) are thus literally the authors of, and responsible for, the state of other parts of nature (human deeds and artifacts). This, surely, is a fact – yet it has an air of paradox, and there is a major philosophical problem about understanding how it can be a fact. The idea of a responsible agent, with the “originative” ability to initiate events in the natural world, does not sit easily with the idea of a natural organism, whose behavior conforms to biological and physical laws under the conditions set by its natural environment and constitution. Our scientific understanding of human behavior seems to be in tension with a presupposition of the ethical stance we adopt toward it: How can something ever be genuinely responsible for what it does, when “what it does” is only an episode in impersonal natural history?

The point of metaphysical tension between natural scientific and ethical perspectives on human behavior has often been identified as the problem of an apparent incompatibility between freedom (or “free will”) and determinism. Responsible behavior must be freely chosen and performed – yet how can it be so if it is causally “determined” by antecedent conditions in accordance with natural laws? I believe, however, that this standard identification is a mistake. As I shall seek to establish in Chapter 1, the perennial and exhausting debate about whether free will is or is not compatible with determinism does have, at its core, an entirely serious basis for doubting whether our ethical self-image coheres with a naturalistic view of ourselves. But as I shall claim, this serious basis for skepticism has nothing essentially to do either with freedom or with determinism.
Rather, it has to do with the possibility of accommodating actions within the natural universe. Persons can be morally responsible only for what occurs through their actions, where an action is here understood (in a semitechnical sense) as an exercise by the agent of his or her powers of control over what occurs. Actions thus seem to involve something foreign to the ontology of natural science: “agent-causation” – the determination of what happens by agents rather than by antecedent events. Thus, the nub of the skeptical problem is that a necessary condition for moral responsibility – the existence of actions understood as episodes of agent-control – appears not to be satisfiable within the causal order as natural science understands it.

This “problem of natural agency,” as I shall call it, would be resolved if it could be shown that the existence of actions (as our ethical perspective understands them) does not, after all, entail the existence of episodes of a special kind of agent-causation. There is a philosophical theory of action (or more properly, a family of theories) that has precisely this consequence. According to the Causal Theory of Action, actions consist in behavior that is caused by appropriate mental states – mental states that make it reasonable for the agent to perform behavior of just that kind. What we think of as agents doing things, it is suggested, is actually a matter of certain of their mental states causing those things to occur. Thus, for example, the theory would maintain that my raising my arm consists in my arm going up as a causal consequence of, say, my intention to raise my arm. On this theory, agent-causation is (in a certain sense) reducible to “ordinary” causation by mental events, and so it fits unproblematically into the ontology of scientific naturalism – provided, of course, that mental events may themselves be understood as consisting in natural, physical occurrences (presumably, of a neurophysiological kind).

The validity of this reduction, however, is far from obvious. In fact, our initial intuitions are strongly to the contrary: Being caused to behave by one’s mental states seems quite different from – and indeed, exclusive of – the kind of action that functions as a necessary condition for moral responsibility. My central project in this essay is to consider whether, despite this initial intuition, a version of the Causal Theory of Action may nevertheless be defended as resolving the problem of natural agency. Is it possible to show that scientific naturalism presents no obstacle to the existence of actions
by establishing that action consists in behavior with appropriate causes of a kind unproblematically admissible within our natural scientific worldview?

My conclusion will be that it is: There is a correct version of the Causal Theory of Action, and the problem of natural agency is thereby solved. The main body of this essay is therefore devoted to developing a causal analysis of action that will survive the various challenges that can be raised against it. I shall seek to blunt, and gradually to erode, the initial intuition that sharply contrasts mentally caused behavior with the kind of genuine action for which the agent might properly be held responsible. In doing so, I shall not, of course, be starting from scratch. Contemporary interest in the Causal Theory of Action stems mainly from Donald Davidson’s defense of reasons as causes of actions in his famous paper, “Actions, Reasons and Causes,”¹ and I shall take as my point of departure a causal analysis of a kind this paper suggests.

My procedure will be to seek successive refinements of this analysis in order, step by step, to overcome the obstacles that the Causal Theory faces. I largely follow Davidson in my view of what these obstacles are: First, there is the challenge posed by the possibility of “weak-willed” (akratic) intentional actions (which I discuss in Chapter 3); and second, there is the more serious problem of causal deviance (which I treat in Chapters 4 and 5). In a word, the problem is to specify the right kind of way in which the agent’s reasons have to cause the appropriate behavior if he or she is to perform a genuine intentional action. Davidson’s own responses to these problems are, of course, of great interest. I shall, in effect, endorse his way of coping with akraisia, although I want to present it in a new way, which draws proper attention to the need to accept intentions as a sui generis class of mental states that can serve as the constitutive causes of action. By way of contrast, I shall repudiate Davidson’s pessimism in the face of the problem of causal deviance: I shall maintain that unless a deviance-excluding causal analysis of action can be produced, there can be no hope of using the Causal Theory to rebut skepticism about natural agency. By critically examining the work of other philosophers sympathetic to the Causal Theory (especially David Lewis and Christopher Peacocke), I seek to develop a causal analysis of action that is immune to deviant counterexamples, and so to support the core ontological

³
claim of the Causal Theory: Actions consist in behavior with suitable mental causes.

Let me now make some more general remarks about the nature, structure, and motivation of my project. First, to avoid a serious (if unlikely) misunderstanding, I should stress that this essay is not a contribution to empirical theories of natural agency. The reader will not find in these pages explanatory hypotheses about how it is that natural organisms put their decisions and intentions into action, nor suggestions about how machines might be constructed to do the same. Such empirical theories belong to neurophysiology, cybernetics, robotics, and cognitive science generally. This essay is, rather, a philosophical prolegomenon to the whole enterprise of developing natural scientific theories of action. It has point only because there are serious grounds for skepticism about the very possibility of a natural organism that is also a genuine agent. If these skeptical arguments are correct, agency will simply fail to be a proper object of natural scientific inquiry. Dealing with these skeptical arguments thus presents a characteristically philosophical task, and it is this task to which my essay contributes.

Although it would be foolish to treat science as irrelevant to philosophical theorizing, the fact remains that there are distinctively philosophical questions. Many of these (perhaps all?) are generated by skeptical reflection on commonsense beliefs and the presuppositions of scientific inquiry — reflection that usually has a relatively a priori character. To put it in concrete terms for the present case, doubts about the existence of genuine agency are doubts about whether actions could exist in the natural causal order as science understands it, not doubts about whether there are any natural actions as a matter of fact. Skepticism here rests on the claim that there is something in the nature of action, as we understand it, that logically precludes its realization in the natural universe — also as we understand it. We may expect to deal with such skepticism, then, only by undertaking an inquiry into our understanding both of agency and of nature. And this inquiry is essentially a conceptual one. It will not be settled by facts about actions in the world — for of course, any such facts will be contested by the skeptic, who doubts whether there can be such facts. If there are facts relevant to settling it, they are not, strictly, facts about actions but rather facts about what we take action to involve, given the way we employ the concept within our ethical perspective on human behavior.
Yet the inquiry cannot be a purely factual one, even granted that the facts sought would have to be facts about our own understanding of action. Although the aim is, of course, to articulate an understanding that already exists among those who use the concept of action (and in the light of that articulation, to judge the force of the skeptical arguments), it is possible that philosophical inquiry will yield some proposals for a refined theoretical understanding of what agency is, and thus serve partly to create the facts about how agency is understood. Our unarticulated commonsense notion of the kind of agency necessary for moral responsibility may well be vague or confused. On examination, skepticism about natural agency might be at least partly traceable to such vaguenesses or confusions – and some conceptual revision may be required to deal with it. Where this proves so, I shall maintain that the intuitions about agency that need revision are neither central nor uncontested.

Second, to remark on the way I have structured my inquiry – on how I understand the burden of proof to fall and why I think a positive defense of the possibility of natural agency is required. My initial stance is reconciliatory naturalism: I start by assuming that the presuppositions of our ethical and natural scientific perspectives are, in fact, mutually consistent. It is tempting to hail this assumption as straightforward common sense. But I suspect that it actually rests finally on a value judgment: that, somehow, it is better for reconciliatory naturalism to be true than for it to turn out either that our belief in agency is mistaken or that, as agents, we belong mysteriously beyond the natural universe that is open to scientific inquiry. (In my own case the roots of this value judgment, I realize with a shock of surprise, lie in attitudes best described as religious.) The skeptical arguments, of course, challenge this stance. To maintain rational belief in reconciliatory naturalism, a reply will be needed to each argument that suggests that systems subject to the laws of nature cannot be originative, morally responsible agents. Each skeptical argument, as it comes up, places the onus on the reconciliatory naturalist – the believer in natural agency – to defend that stance.

None of the skeptical arguments, I believe, is decisive. Nevertheless, some of them have considerable plausibility. They call up antinaturalist intuitions about agency, which, I shall claim, can be dismissed only in the light of a positive case for natural agency – a
theory of agency that decisively shows that actions have a place in our natural ontology and, therefore, that skepticism about natural agency must be misplaced. I believe that it is the Causal Theory of Action that offers the best prospects for a positive case of this kind, and in Chapter 2, I seek to establish its superiority over alternative approaches. I reject any attempt (such as made recently by Daniel Dennett) to reconcile the ethical with the natural perspective “on the cheap” by adopting an antirealist account of action. And I am also suspicious of libertarians, such as Peter van Inwagen, who hold that the reconciliation can be achieved — but only because the natural universe is not a deterministic system. I shall argue that a successful Causal Theory of Action will be enough to vindicate the natural possibility of action, whether the universe is deterministic or not.

Finally, a comment on my motives: Recent philosophical literature is rich in discussions of agency, and in particular, of the merits or otherwise of the Causal Theory of Action. There is some virtue, I believe, in attempting a unified critical account of the current state of these discussions. To a degree, my motivation in this essay is to make a modest attempt along these lines (especially with respect to the state of the art on the causal deviance problem). But more important, I wish to situate some of the increasingly abstruse topics beloved of “action theorists” in a context that reveals their full significance.

Many philosophers proceed as if the topic of agency were somehow sui generis — an intrinsically given item on the philosophical agenda. Adequate explanations of what is really philosophically problematic about agency are surprisingly rare. In this essay, I seek to provide such an explanation by relating theories of agency to the problem of freedom and determinism. As I have said, it is my view that this problem is, at root, a problem about the very possibility of agency in the natural causal order — a problem that can be solved by successfully defending the Causal Theory of Action. Although recent debate on freedom and determinism has been extensive and vigorous, the significance for it of the correctness or otherwise of the Causal Theory of Action has not been properly emphasized. I aim to repair that omission. I hope to show that what keeps the old freedom and determinism debate alive (in the face, especially, of scientific rejection of the deterministic hypothesis) is its relation to the problem of natural agency — and that is the context which gives
philosophical discussion of agency (and especially of the Causal Theory) its central point.

Davidson, whose work on the Causal Theory of Action has been
the greatest single philosophical influence on this essay, has ex-
licitly recognized the significance of the Theory for the problem
of natural agency. He claims that the Causal Theory “is enough . . .
to explain the possibility of autonomous action in a world of
causality.” Yet Davidson does little to explain why this should be
so and shows no patience for debating with incompatibilists (who
hold free action to be impossible under determinism). Furthermore, he finds himself, because of the causal deviance problem,
unable to provide a causal analysis of action that would positively
prove the truth of a Causal Theory capable of playing this vital
philosophical role. I hope to repair these unfortunate gaps in
Davidson’s position. I want to establish that a Causal Theory of
Action is just what we need to defend the possibility of natural
agency (though, as I shall mention shortly, autonomy may be
another matter). And I want to show that a Davidson-inspired
causal analysis can, despite its inspirer’s pessimism, be brought to
proper completion. It should thus be clear that the detailed con-
sideration I shall give to the problem of causal deviance is no
isolated academic game. As I see it, dealing with causal deviance is
actually at the cutting edge of the attempt to rebut a historically and
currently influential source of skepticism about the place of persons
in nature.

That said, however, I should make it clear that I do not profess to
be dealing here with the whole philosophical problem of how it is
that persons can exist in the natural world as conceived by science.
My concern focuses on one aspect of personal existence: the fact
that persons are agents, in the sense that they exercise their own
control over outcomes. I acknowledge that skepticism about mak-
ing intelligible the place of the person in nature might be based on
other features of personhood besides this one. There seem to be, in
fact, two other aspects of personhood that might present difficulty
for the naturalist—one, so to speak, on either side of the problem of
natural agency.

There is a problem prior to the problem of natural agency—
namely, the problem of how conscious minds can be part of the
natural order. Surely we may understand how agency is naturally
possible only if we first understand how mentality may be part of
nature? That this is so is entirely clear if a Causal Theory of Action is to provide the solution to the problem of natural agency because this theory holds that action consists in behavior caused by relevant mental states. And there is a problem posterior to the problem of natural agency — namely, the problem of explaining how those extra properties beyond agency as such that are required for personal moral responsibility can themselves be realized within a natural scientific ontology. Although capacity for agency is (as I have been stressing) a necessary condition for moral responsibility, it is only a necessary and not also a sufficient condition for it since there can be beings who can exercise their control in action, yet who cannot appropriately be held morally responsible for the way they do so. Young children and some nonhuman animals provide cases in point.

So there could in theory be three grades of skepticism about the possibility of making intelligible the place of persons in nature as science understands it. First, there is skepticism about understanding how minds (or, better, beings capable of mental states and functions) can be part of nature. This, of course, is the mind/body problem. Second, there is the skepticism that accepts the natural possibility of minds, but doubts the possibility of a coherent naturalist account of agency, understood as the capacity to exercise one’s own control over outcomes. This is the problem of natural agency. Third, and finally, someone who has solutions to both the mind/body problem and to the problem of natural agency might still persist in doubt because he or she believes that those extra properties required for an agent to be a moral agent cannot be understood as realizable within our natural ontology. Without wishing to be precise about what these extra properties are, I shall simply dub this the problem of natural moral autonomy. Thus, to vindicate reconciliatory naturalism completely — to show that our ethical perspective coheres with scientific naturalism — one would need to rebut skepticism at each of these points.

In this essay, I do not attempt a complete vindication of reconciliatory naturalism. I seek to deal only with the second of the three skeptical problems just outlined. I simply assume that the first problem can be solved — that a naturalist account can be provided of mental states and how they can function as causes. Under that assumption, I then consider whether the existence of agency poses any further problem for a naturalistic viewpoint. As to the third problem, the problem of natural moral autonomy, my view is that
it is far less serious than the other two. Those who have doubted whether moral agency can be realized in the natural order have, I believe, been puzzled essentially about how nature as science understands it could accommodate systems with a genuinely origina-
tive power of control, giving them the capacity to act for reasons of their own in shaping the course of natural events. And this, of course, is just the problem of natural agency. Once you grant the natural possibility of agency, I believe that all you then need for full moral autonomy are certain extra mental capacities whose natural possibility does not present any new category of metaphysical problem. I believe, then, that resolving skepticism about natural agency does remove the metaphysical doubts we have about whether we really can see ourselves the way we want to – as morally responsible persons who belong, entire and complete, to the order of nature as our science understands it. As I shall explain in Chapter 6, however, this position is not equivalent to a full naturalist justification of the rationality of our ethical practice in assigning responsibility.

NOTES

5 Davidson, Essays on Actions and Events, p. 88.
6 See ibid., p. 63.
1

The problem of natural agency

I. A theory in search of its problem

This essay is an inquiry into a philosophical theory of agency that claims that actions are events with a special type of causal history. To put it briefly, according to this Causal Theory of Action, to act is to be caused to behave by mental states of one’s own – mental states that make the behavior reasonable in the circumstances. But what is the point of such a theory, and why is it worthwhile to conduct an inquiry into its truth? These are the questions I shall address in this chapter. As with all philosophical theories, the point of the Causal Theory of Action may be understood only in relation to the philosophical problem it professes to solve. What, then, is the problem that the Theory claims to resolve, and why is it philosophically important? Starting from my dissatisfaction with a standard answer to this question, I shall argue that the Causal Theory of Action offers an answer to one of the major perennial metaphysical problems – the problem of “freedom and determinism.” The quotation marks are used advisedly because, as I shall argue, the problem essentially concerns neither freedom nor determinism, but is rather about making intelligible the possibility of agency within the natural order. And this problem is one that must be solved by anyone who seeks to defend a fundamentally naturalistic account of human personhood. Thus, this chapter will present the case for taking the Causal Theory of Action seriously on the grounds that if it is true, it resolves what I shall call the problem of natural agency, and thus advances the naturalistic program for understanding the place of human persons in the universe.