

INDETERMINISM AND FREE AGENCY: THREE RECENT VIEWS

Timothy O'Connor

Philosophy and Phenomenological Research, 53 (1993), 499-526.

I

It is a commonplace of philosophy that the notion of free will is a hard nut to crack. A simple, compelling argument can be made to show that behavior for which an agent is morally responsible cannot be the outcome of prior determining causal factors.¹ Yet the smug satisfaction with which we incompatibilists are prone to trot out this argument has a tendency to turn to embarrassment when we're asked to explain just how it is that morally responsible action might obtain under the assumption of indeterminism. Despair over the prospect of giving a satisfactory answer to this question has led some contemporary philosophers to a position rarely, if ever, held in the history of philosophy: free, responsible action is an incoherent concept.²

In recent years, as the enterprise of speculative metaphysics has attained a newfound measure of respectability, incompatibilist philosophers who are inclined to think that freedom of action is not only possible, but actual, have re-emerged to take on the formidable task of providing a satisfactory indeterministic account of the connections among an agent's freedom to do otherwise, her reasons, and her control over her action. In this paper, I want to examine three of these proposals, all of which give novel twists to familiar themes.³ I will argue that despite the considerable ingenuity these philosophers evince, their attempts do not succeed. A common criticism of these theories will be that they fail to give a satisfactory account of what I term "agent-control", a certain feature of actions whose presence I take to be a central requirement for any workable model of responsible agency. I believe that the general notion I try to capture under this label is implicit in much of the voluminous discussion of the problem of free will (especially in compatibilist criticisms of libertarianism), although I am unaware of any explicit formulations of it in just the way I have in mind. Simply put, agent-control is that feature of the process of agency that accounts for how a particular piece of behavior is connected to, or an 'outflowing of', the agent, i.e., that which allows us properly to assert that the action was controlled by the agent.

Unfortunately, this rather schematic way of spelling out the concept is not especially informative, but if I were to be any more specific than that I should have to draw upon some particular model of agency, and I think it is more useful to note first the more general feature that is common ground to many theories, compatibilist and

incompatibilist alike. The polemical purposes of these two factions have often kept the participants in the debate from clearly identifying those general requirements on responsible agency upon which all, perhaps, can agree.⁴ Partly as a consequence of this, philosophers often do not properly disentangle closely related issues when attacking the account offered by an opponent.

To return to the matter of 'agent-control' (as I intend to use this term), compatibilist theories of action of the standard sort clearly do imply such a feature. For on these accounts, an intentional action of an agent is a causal consequence of the (dominant) beliefs and desires she had at the time of her action. On such theories, then, agent-control is realized through an external, causal relation between her character at that time (in a broad sense, including relatively momentary attitudes and beliefs) and the action itself. (But incompatibilist approaches will attempt to account for this feature in alternative ways.) As an incompatibilist, however, I do not believe that any compatibilist account of human agency can allow for there being alternative courses of action that are genuinely open to the agent. Here, too, many compatibilists will accept the availability of alternative possibilities as a further requirement on responsible agency, though disagreeing on what satisfaction of this condition comes to.

Virtually all incompatibilist theories of action, I believe (and certainly the three that are discussed in this chapter), satisfy this second requirement in positing significant causal indeterminism in the triggering of action, in such a way that very different actions are causally possible in those circumstances. But, I will argue, not all such theories allow for this (given their account of agent-control) in a way that explains how it could be up to the agent which alternative is realized. That is why I said earlier that I will argue that none of the theories discussed in this chapter gives a satisfactory account of agent-control. For while my criticism of one of the theories will be that, on it, actions lack the property of agent-control altogether, my criticism of the others will begin by conceding that they provide a way in which agents both control their actions and have alternative courses of action open to them, and will then go on to insist that (nonetheless) they fail to do so in a way that gives the agent a choice about which of the possible actions is in fact undertaken. My position, thus, is that responsible agency requires not only the satisfaction of the two separate conditions of agent control and alternative possibilities, but also the proper confluence of these factors as they are realized within the theory.⁵

One libertarian theory that does seem - at least on the face of it - to satisfy the constraint of agent-control in an appropriate way (i.e., that achieves the confluence of which I have just spoken) is variously referred to as "agent causality" or "the agency theory". This approach, though, has problems of its own, problems that many have

thought fatal to the theory. Each of the theories I discuss here attempts to make do without the seemingly troublesome claims of the agency theory. Indeed, they are representative of the two⁶ basic alternatives to the agency theory for the libertarian.

While my aim in this paper is thus largely a negative one, it is no part of my purpose to contribute to the frankly tiresome trail of negative pronouncements concerning the prospects of giving a viable indeterminist account of free agency. Rather, it is more the ground-clearing phase of a more constructive project. For I believe, contrary to the firmly-entrenched received view, that the concept of agent causation is not dark and impenetrable and it can, furthermore, (for all we currently know) be instantiated in a world of quarks, photons, leptons, and gluons. But making good on these bold claims will have to be reserved for another occasion.⁷ Here I must rest content with making a case to my fellow incompatibilists that an acceptable libertarian theory, if one is to be had at all, must lie in the direction of the beleaguered agency theory.

II

The first proposal of an indeterministic account of free agency that I want to consider is the one suggested in Ginet (1990). Ginet's account is a version of what is sometimes called "simple indeterminism", the general view that free agency doesn't require there to be any sort of causal connection (even of an indeterministic variety) between the agent and his free actions.

The central feature of Ginet's theory of action is the claim that there is a simple mental action (lacking internal causal structure) at the core of every causally-complex action (p.15). In some cases, such simple mental actions are complete in themselves, not constituting part of a more complex action, as in mentally saying a word or forming an image. Simple mental acts of this sort differ from passive mental occurrences such as unbidden thoughts or images intrinsically, in virtue of what Ginet terms an "actish phenomenal quality" - the quality one might characterize (in the case of mentally saying a word) as "it's seeming as if I directly produce the sound in my 'mind's ear'." Ginet emphasizes that this quality "belongs to the manner in which the word occurs in my mind and is not a distinct phenomenon that precedes or accompanies the occurrence of the word" (p.13).

In the central sort of case, however, this simple action will be a volition that constitutes the core element of one's voluntarily exerting the body, causally producing that exertion (pp.23ff.). The intentional content of the volition is that you are voluntarily exerting your body in a specific sort of way. This content, though, is directed to the

immediate present (thereby differing from decision or intention), making it a "fluid" sort of mental activity over time (p.32).

To take a very simple example, my raising my finger is thought to consist of (a) a continuous sequence or "flow" of volitions, each of which has the content that I exert force of a certain degree in a certain direction with that finger (and has the phenomenal quality of seeming as if I 'directly' make it occur), and (b) the movement of the finger being appropriately causally connected to these volitions.

In this brief account of Ginet's basic picture of action, I have left out many interesting details, but I think that this will be sufficient background for understanding how he addresses the problem of control that seems to arise when we add the assumption that, in ordinary (or at any rate, some) cases, the simple mental actions that constitute volitions are themselves causally undetermined. Those who are inclined to see such a problem, Ginet suggests, may be thought of as accepting the following principle:

(Q) For any time t and any undetermined event occurring at t : It is not possible for it to have been open to anyone to determine whether that event or some alternative (undetermined) event instead would occur at t .

But, he argues, to hold that Q applies without restriction, extending even to the very actions of an agent, is a mistake:

To determine an event is to act in such a way that one's action makes it the case that the event occurs. Let us grant (for the sake of this discussion) that if the event is not one's own action, then this requires that the event be causally necessitated by one's action (in concert with other circumstances) and thus that it not be an undetermined event. But if the event is one's own action, then one's determining it requires only that one perform it; and one's performing it, which is just the action's occurring, is compatible with that event's being undetermined, not causally necessitated by antecedents. (p.127)

In evaluating this response, it should be noted first that the latter claim about what is required for "determining" an event internal to oneself - that it simply occur - is restricted to actions alone, and second that an event is an action in virtue of intrinsic properties of that event, quite apart from any external causal relations that obtain between that event and either other events internal to the agent or the agent himself. So, for example, I may not be said to determine that an episode of my involuntary blinking occurs, whether or

not there happened to be a certain amount of indeterminism in the series of micro-events underlying this occurrence. And the reason for this resides wholly in the fact that the occurrence lacks the characteristic fluid mental activity of its seeming to me as if I directly bring about this particular exertion of my body (and thereby lacks as well a (non-deviant) causal connection between such mental activity and the physical events comprising the blinking episode).

This proposal seems to me unsatisfactory. It is, first of all, unclear what content the notion of "determining an event" is thought to have here.⁸ One might suppose it to be causal in nature (similar to "bringing it about that"), but this is clearly not the case if it can apply in virtue of an event's having certain intrinsic properties. But if not causal, then what? It might be suggested that it simply signifies that an agent bears one of two sorts of relations to an event, each appropriate to one of two different classes of events - our actions and their (immediate) causal consequences. But this seems implausible given the radically different type of relations we bear to each class - purely internal on the one hand, and external and causal on the other - on the proposal under consideration.

I suggest as a more direct criticism that Ginet is able to purchase some plausibility for his claim (that we may be said to determine our own action simply in virtue of its obtaining) by, in effect, taking the reader through the following procedure: first, one considers a case which has the usual introspective quality that essentially forms the basis for the supposition that one has "determined" one's own behavior. Then, we note that it is quite consistent with this that the volitional events lying at the core of our behavior are causally undetermined. The conclusion we are asked to draw is that it is in no way inconsistent with that supposition that our volitions are causally undetermined. I think, however, that reflection ought to show us that this approach to settling the matter is unsound, and that the subjective plausibility of Ginet's claim diminishes when we take note of this fact.

To take a simple case, then, we may consider an undetermined simple mental action (a volition to move my right arm in a particular way, say). This volitional event has the right intrinsic quality - its seeming as if I directly brought it about - to qualify as an action on Ginet's account. It is not entirely clear to me whether our pre-theoretical concept of action permits this to count as a sufficient condition for deeming a mental event (or connected sequence of events of which it forms the initial segment) an "action", but the question of whether further, causal presuppositions of a certain sort enter into this notion is difficult. One who is inclined to take issue with Ginet here will probably do so on the same sort of grounds on which I am challenging his analysis of what is required in order to

determine one's action. Hence, forcing the issue in this way will result only in shifting slightly the terms of the dispute.

So let us allow that the (undetermined) volition we are supposing to have occurred is an action. Should we allow that I determined it to occur? I agree that we would naturally suppose this when attending only to the character of the experience, which is usually all we have to go on in actual practice. Since it seems to me that I have (in some sense) directly brought this event about - it has, as Ginet puts it, the familiar "actish" phenomenal quality, or, in an older terminology, is the (apparent) experience of an exertion of active power - I will no doubt presume that I have determined it. But, I would claim, we recognize that such judgments formed on the basis of the way in which our behavior is experienced are, in principle, defeasible. For they are based, in part, on our assumption that things are pretty much as they seem with respect to our own behavior, and hence that its seeming as if I directly bring about my own volitions (or, if one is doubtful concerning these, my own decisions) is owing to the fact that I do so bring it about. But if I'm given strong reasons to believe that this is not actually the case, then I would withdraw the claim that I determined it. (In this respect, perhaps, the concept of determining one's own action may be stronger than that of acting itself.) Hence, *pace* Ginet, the causal ancestry of such events is relevant to the matter under consideration.

A further reason to suppose that Ginet's analysis of what it is to determine an event is inadequate arises when we bring on stage Harry Frankfurt's infamous neurophysiological manipulator, Dr. Black. If Black directly causes within me a volition to move my arm in a specific sort of way, Ginet's account seems to say that I may nonetheless be said to have determined it. But this, I believe, is plainly wrong. So perhaps we should amend Ginet's account so that one's determining one's action requires only that one perform it and that it not be determined by anything else.

The case of Dr. Black underlines the fact that the notion of "determining" an event clearly involves some notion of autonomous control. Is this type of control that we take ourselves to exercise over our own actions (which I earlier called "agent-control"), itself fundamentally causal in nature? Ginet would claim that it isn't. On his view, we do control our bodily exertions insofar as they are appropriately caused by our volitions. So the limited control we have over our overt bodily movements - and through them, other events in the world - is a direct result of our capacity to will efficaciously. Yet, as Ginet recognizes, these volitions are themselves simple actions we perform. And these, he suggests, need not be caused for us to control them. But how can it plausibly be maintained that the fundamental actions in virtue of which we have control over events within and immediately external to our body may themselves be simply uncaused

occurrences; that we may be said to be in control of these simply by virtue of their obtaining (undeterminedly)? Perhaps the challenge is most directly posed in these terms: why are volitions intrinsically such as to confer control in the absence of determining antecedents?

Consider again the involuntary blinking episode. This event lacks (we are supposing) the experience of its-seeming-to-me-as-if-I-bring-it-about that is characteristic of voluntary activity. So it seems clear (as Ginet's account implies) that this event is not properly described as an action of mine, nor does it seem that I have control over its occurrence. Compare this with a case in which my blinking does have as a core component the volition to do so, and suppose this volition to be causally undetermined. This second case does describe an action, on Ginet's account. Furthermore, it is an event over which I am said to have control, in virtue of the volitional core action. But again, why should we suppose this to be so? Why does the addition of a causally undetermined volition, in and of itself, suffice to transform the first (involuntary) behavior into one that I control?

So far as I can see, Ginet's discussion, original as it is in defense of a type of simple indeterminist model of libertarian agency, does not point the way towards answering that question, nor does it give a reason to suppose that the question is somehow improperly posed. His discussion has, though, (to my mind) clarified the precise point of contention between the simple indeterminacy theorist and her critics: whether or not control over one's actions obtains in virtue of certain purely intrinsic, non-relational features of those actions together with the negative condition of there being no determining prior causes. I am inclined to think that the onus is clearly on one who affirms this thesis to explain why we should think this to be so. Hence, given the absence of positive support for what seems to be a counterintuitive thesis, I conclude that agent-control, as with the control we exercise over events other than simple mental events, can obtain only via some sort of causal relation between ourselves and our actions. The compatibilist maintains that such a relation is realized through a process in which the agent's beliefs and desires prior to and simultaneous with her volitions and/or decisions are the primary factors in causally necessitating that volition/decision.⁹ And the agency theorist posits a more direct causal relation between agent and action, which I will not attempt to explicate here.

III

It might be suggested that the indeterminist can follow the compatibilist's basic strategy by suggesting that the agent controls her actions through her prior reasons. Such reasons (perhaps conceived as belief-desire complexes, where the term "desire" is

construed quite broadly) may be thought to cause, but not necessitate, the action. Some would assert that the very concept of causation involves the notion of necessitation, and so deny that the proposal is a coherent one.¹⁰ For my discussion here, I will simply ignore such an objection. Even granting this assumption (as I am inclined to do anyways), this sort of approach (which I will term "causal indeterminism") is, I believe, faced with a serious difficulty. In this section of the paper, I will try to set out this difficulty in generic terms. In subsequent sections, I will address two particular causal indeterminist theories of free agency that attempt to overcome it, and I will argue that they fail to do so.

One way of articulating the general problem is set out by van Inwagen (1983).¹¹ He discusses the case of a thief confronted with an opportunity to rob a poor-box, and having reasons not to take the money as well as obvious incentives to do so. After a brief deliberation, he repents from his theft in progress and walks away. We will further assume, in accordance with the proposal presently under consideration, that his act of repentance ("R") was caused but not necessitated by the relevant belief-desire complex ("DB") constituting his prior motivation to act in that way. Van Inwagen writes:

But note that DB did not have to cause R. Moreover, since DB did not have to cause R, and since DB alone caused R, R did not have to occur. But then did the thief have any choice about whether R would occur? [It would seem not.] ... Once DB has occurred, then everything relevant to the question whether R is going to happen has occurred. After that we can only wait and see. In a perfectly good sense, it is going to be a matter of chance whether R occurs, whatever sophisticated difficulties some philosophers may raise about defining this notion.

He then summarizes this conclusion in terms of the following principle:

(S) If an agent's act was caused but not determined by his prior inner state, and if nothing besides that inner state was causally relevant to the agent's act, then that agent had no choice about whether that inner state was followed by that act.

Interpreting his notion of "having a choice about" in terms of my notion of agent-control, one might re-state his reasoning thus:

If it was an open, non-negligible possibility, causally speaking, that a different act have followed the antecedent circumstances, and if having control over one's act must involve a causal capacity, as argued earlier, then it seems that one's control

diminishes in direct proportion to the prior likelihood that the effect not have resulted. The problem may be put thus: the suggestion under discussion subscribes to the general model of agent-control given by standard (deterministic) compatibilist analyses. But this can be used to generate an indeterministic account of free agency only by weakening the very (causal) relation in virtue of which control is deemed to obtain on that model.

Bishop (1989) does read van Inwagen this way and argues that this line of reasoning is not cogent.¹² In remarks partly reminiscent of Ginet, he suggests that

we are bound to be seriously misled if we interpret a causal theory of an agent's direct exercise of control [i.e., control over one's own actions] by using a model of indirect agent-control over an outcome - and this applies whether our [causal theory of action] is probabilistic or deterministic. (p.71)

For, he continues, we indirectly control events external to us by performing actions that, in the circumstances, cause those events, whereas we do not directly control our actions themselves by performing other actions. Rather, direct control of actions is constituted by causal relations between mental states and behavior which, according to the indeterminist, are only probabilistic (p.72).

The point of Bishop's remarks seems to be to suggest that the critic of a causal theory of action in which control obtains in virtue of merely probabilistic causal relations is guilty of the following confusion: He is (perhaps unconsciously) thinking of the first element of the posited relation as a further "doing", and so (correctly) inferring that control varies directly with the probability that the second element will occur, given this "doing". But of course the causal relation does not link two such actions, but rather a set of prior mental states and an action. We do not, to repeat, control our own actions by performing other actions that result in them; instead, control obtains simply in virtue of a causal relation between our "inner state" and action. And it may perfectly well be the case that this causal relation in virtue of which control obtains is probabilistic only.

Bishop is certainly correct in noting that if a causal relation of the appropriate sort can constitute a form of agent-control (as causal theories of action in general maintain), there is no clear reason why that relation cannot be non-necessitating (assuming non-necessitating causation to be possible). The point should be fairly evident if we bear in mind the reason one might have for supposing that a deterministic causal theory of action provides a viable form of agent-control. Our desires and beliefs, particularly those that are

relatively stable over time, are a central feature of our conception of ourselves. So for our actions to match these states (in an appropriate fashion) and be causally produced by them is for such actions to be a product of "who we are" (at the time they occur). But none of this requires that the actions be a deterministic outcome of prior mental states, only a causal outcome. So a criticism of this proposal for modeling libertarian free agency on the grounds that the causal relation is less than deterministic seems ill-conceived.

And once this is recognized, it may seem that this proposal provides just what is needed. Deterministic causal theories can underwrite a form of agent-control, but fail (as the libertarian sees it) to allow for genuine alternative possibilities open to the agent, and therefore fail to depict free, responsible agency. And the simple indeterminist approach (of which Ginet was our representative author) yields genuine alternative possibilities without (I have argued) agent-control. The causal indeterminism now being considered allows for both. It suggests that in many circumstances, persons have (at least partially) distinct desire-belief complexes (more simply, "reasons") that point toward different courses of action, and that (in accordance with libertarianism) the performance of any of these - not just the one that was actually done - would "graft" coherently onto precisely the same prior circumstance. Each such reason is a potential cause of the corresponding behavior; whichever action is (as it happens) undertaken will have been caused by the matching reason(s). We cannot, of course, specify with certainty which reason will function as cause prior to the action itself, but this perfectly accords with the libertarian contention that (free) actions are determined only at the time of their occurrence.¹³ Finally, in no case does an action occur that is simply uncaused. Have we, then, a solution to our problem?

Unfortunately not, I believe. And I think that a closer examination of van Inwagen's discussion suggests the nature of the problem.¹⁴ In broad terms, it seems doubtful that the form of agent-control adopted - one constituted by a causal relation between prior mental states and action - operates in a manner consistent with the general picture of free agency that motivates libertarianism.¹⁵ And this is perhaps not surprising, since it is an only slightly modified version of a conception of agent-control devised explicitly to elaborate the very different picture of free agency put forth by compatibilism. For even though the causal indeterminist account (as sketched in the previous two paragraphs) allows for the real possibility of different courses of action, any of which would be "controlled" by the agent in the sense that it would be the causal product of the agent's own reasons, it doesn't seem to be 'up to the agent', something he 'has a choice about', to determine just which potential cause will be efficacious in any given instance, and so which action will actually occur. It would seem, rather, that it is simply a matter of its

falling under a statistical or quasi-statistical tendency governing the general pattern of behavior in types of circumstance over time, and this probabilistic tendency clearly is not something the agent has any choice about.

As van Inwagen suggests, the scenario seems to be this. On any given occasion,

there are two events its coming to pass that DB causes R and its coming to pass that DB* causes R* such that, though one of them must happen, it's causally undetermined which will happen; and it will have to be the case that the thief has a choice about which of them will happen.

He appends to this the fitting comment that

[i]f this were so, I should find it very puzzling and I should be at a loss to give an account of it.¹⁶

Now one might try to respond to this¹⁷ by questioning the claim that the causal indeterminist needs to hold that the thief has a choice about which of the two events van Inwagen describes will happen. Why not simply insist that what the thief has a choice about is which action to perform (R or R*), not which reasons will cause that action (even if DB causing R is causally necessary for R)?

In order to evaluate this suggestion, let us be clear about our locution, "having a choice about". It will be agreed on all sides (compatibilists and incompatibilists alike) that this property is not instantiated solely in virtue of the occurrence of a mental event of the sort we denominate "decisions". For were Dr. Black directly to cause such an event within me, bypassing my deliberative capacities, then even the compatibilist will agree that though (perhaps) one might say that I chose or decided to act as I did, I certainly didn't 'have a choice about' doing so. It simply wasn't 'up to me', in virtue of its disconnectedness from my beliefs and desires. So it seems that where the property of having a choice about X is instantiated, the agent will have freely chosen to do X (or the obtaining of X will be a consequence of some free choice of the agent).

Now the suggestion we are presently considering is that while (1) DB obtains, and the thief, of course, doesn't at that time have a choice about whether DB obtains¹⁸, and (2) (DB \rightarrow R) is true, and the thief doesn't have a choice about whether (DB \rightarrow R), the thief nonetheless does have a choice about whether R obtains. This, however, implies the invalidity of the following rule of inference:

$N_{st}(p), N_{st}(p \rightarrow q), \vdash N_{st}(q),$

where N is a propositional-operator form that is to be interpreted as 'p and agent s did not have a choice at time t about whether p' (and the arrow represents the material conditional). The causal indeterminist is hardly in a position to reject this principle, since it is precisely what underlies the usual argument for incompatibilism!¹⁹ While I cannot defend this claim here, I think that it is very doubtful whether one could (reasonably) reject this principle and still remain an incompatibilist.

Therefore, I will assume in what follows that the incompatibilist who wishes to develop the causal indeterminist approach must confront directly the challenge of explaining how it is that an agent can have a choice about which potential indeterministic (reasons-) cause will be efficacious in a given situation.

IV

Having set out the general problem facing causal indeterminist theories, I will now turn to the remaining two theories of free agency to be discussed in this paper, each of which are variations on the basic causal indeterminist picture and which attempt to deal directly with this problem. I begin with the account offered by Robert Nozick.²⁰ Our challenge to Nozick, then, is to show how an agent's action might be caused (but not determined) by (one or more of) her reasons in such a way that she is responsible for the fact that that particular action was causally produced by those reasons rather than an alternative action's having been caused by some other set of reasons.

The first part of Nozick's theory is an account of the process of decision:

reasons do not come with previously given precisely specified weights; the decision process is not one of discovering such precise weights but of assigning them. The process not only weighs reasons, it (also) weights them.

... It is neither necessary nor appropriate, on this view, to say the person's action is uncaused. As the person is deciding, mulling over reasons R_a which are reasons for doing act A and reasons R_b which are reasons for doing act B, it is undetermined which act he will do. In that very situation, he could do A and he could do B. He decides, let us suppose, to do act A. It then will be true that he was caused to do act A by (accepting) R_a . However, had he decided to do act B, it then would have been R_b that caused him to do B.

...The weights of reasons are inchoate until the decision. The decision need not bestow exact quantities, though, only make some reasons come to outweigh

others. A decision establishes inequalities in weight, even if not precise weights. (pp.294, 295, 296-7)

Nozick goes on to suggest that the orthodox interpretation of the quantum mechanical theory of measurement provides a structural analogue to his analysis of decision as assigning comparative weights to previously unranked reasons. On this theory, a quantum mechanical system is in a continuously-changing "superposition", or probability mixture, of states, which changes discontinuously at the time of a measurement. The measurement is said to "collapse the wave packet", reducing the superposition to a particular state in an undetermined fashion. Analogously, on Nozick's account, a person has reasons with unspecified weights prior to decision (and so may be said to be in a superposition of weights with respect to his reasons). The process of decision reduces the superposition to at least a determinate comparative ranking, and it is not determined to which such ranking it will be reduced. (p.298)

It is unclear how generally Nozick takes this account to apply to the decision-making process. He does say that assigned weights do not disappear after a decision, but instead set up a tentative framework for (relevant) future decisions. Some episodes of decision in the face of conflicting reasons, then, may simply make use of a framework of weights established earlier. He maintains, however, that this is always subject to being adjusted or overturned (a move apparently designed to safeguard the freedom of such decisions), and suggests that we might take this to imply (on his theory) that "there is always present a reason of indeterminate weight to reexamine and overturn an earlier precedent, which reason itself must be given a determinate lesser weight in the decision to follow the precedent" (p.298). It seems, then, that Nozick would rule out the possibility that an agent perform an action that he deems to have less weighty support by reasons than another action available to him.

Given this general theory of the decision-making process, we must adjust slightly the terms of the challenge posed at the end of section III. Which reason the agent will act upon is undetermined prior to the decision. But that action is causally necessitated by the immediately prior, reasons-weighting decision. (Or perhaps we should think of the reasons-weighting decision as the initial, originating segment of the larger action.) Our question, then, is this: how does Nozick reply to the claim that it is an arbitrary, random matter what character the reasons-weighting decision will have, and so it is not in any clear sense up to the agent what that decision will be? To find a satisfactory answer, he suggests that we consider

the policy of choosing so as to track bestness: if the act weren't best you wouldn't do it, while if it were best you would. The decision to follow this policy may itself be an instance of it, subsumed under it.

In other words, deciding to adopt the policy of tracking "bestness" may be the best general action-guiding policy one might adopt, and so the decision to do so is itself in accordance with that very policy. Similarly, many decisions to act in a specific way may be self-subsuming insofar as they bestow

weights to reasons on the basis of a then chosen conception of oneself and one's appropriate life, a conception that includes bestowing those weights and choosing that conception (where the weights also yield choosing that self-conception). Such a self-subsuming decision will not be a random brute fact; it will be explained as an instance of the very conception and weights chosen... It will no more be a random brute fact than is the holding of a fundamental deep explanatory law that subsumes and thereby explains itself. (pp.300-1, emphasis added)²¹

The reader may well be surprised by the postulation that the fundamental law(s) of nature, whatever it may be, is explanatory of itself in virtue of its being subsumed under itself. Nozick (in a separate discussion of this matter) is driven to this suggestion by the fact that there appear to be only two alternatives - such laws are either themselves unexplained, brute facts or necessary truths - and he believes that neither of these are satisfactory. It is not at all clear to me, however, that his proposal, while highly imaginative, is a coherent one. Suppose, he writes, that the following principle P is the most fundamental law of nature:

P: any lawlike statement having characteristics C is true.

The sort of characteristics Nozick has in mind here are purely formal ones. (He mentions invariance and symmetry.) That true lawlike statements can be identified by such characteristics alone is highly dubious, but we can simply ignore that. Now, Nozick claims, if P itself is a lawlike statement having characteristics C, then given that it is true, the reason why may be "the very content that it states" (p.119) - i.e., in virtue of its having characteristics C.

One might suspect that there are completely general objections to this notion of explanatory self-subsumption, but I think there is a more direct criticism of its

application in this particular case: it seems totally implausible to suppose that any statement along the lines of P could itself bear all the characteristics it specifies. For these are criteria intended to characterize statements linking patterns of events, and so they seem inapplicable to P itself, which has a very different form. Nozick does not address this obvious objection, so far as I can see. So we have yet to be given much basis for confidence in there being so much as a possibility of instances of the unique form of explanation he postulates.

But let us leave this general issue and return to the application that is of interest to us here. Suppose we have a decision in which the weights it bestows "fix general principles that mandate not only the relevant act but also the bestowing of those (or similar) weights." Can we explain how it is up to the agent that those weights (rather than some others) are assigned, by noting that the decision is "an instance of the very conception and weights chosen"?²²

I cannot see how we can. We may put aside the (considerable) doubts one might raise about the possibility of an explanatory law that explains itself by being self-subsuming, and suppose this presents a "live" option. But this is of no help to us here, because of the disanalogy between an explanatory law and a decision that institutes a general policy or conception in accordance with which one is to act. Explanatory laws do not become true at some moment in time. If there were an analogy here, it would have to be between self-subsumptive decisions and an event of a law's coming to be true, which event was explained by, because subsumed under, the very content of the law. (A suggestion of this sort would sound perilously like a rather bad version of the ontological argument sometimes attributed to Descartes.²³) But if the analogy with putatively self-subsumptive explanatory laws breaks down, then there doesn't seem to be a basis for claiming that a decision can be explained (solely) by reference to its own content.²⁴ What sort of subsumptive relationship does the content of a decision of the sort Nozick describes bear to the decision itself, if not a relationship of explanation? Simply one of sanctioning, prescribing or justifying that decision. If I come to order certain values (whether highly narrow and circumstantial or quite broad, extending to a general conception of my life) during a decision-making process, this ordering subsequent to the decision will "affirm" that decision precisely because the latter is an act that is in accordance with it.²⁵

I conclude, then, that Nozick has failed to provide a tenable solution to the problem posed at the end of the previous section.

To summarize the results obtained thus far: A viable theory of the nature of responsible choice-formation seems to require the proper confluence of two factors:

alternative possibilities and agent-control. It appears that simple indeterminist theories fail to give an account of agent-control, an important condition on free agency. (This was argued via an examination of the account recently provided by Ginet.) Causal indeterminist approaches do give some sort of account of this, but it is of a rather weak sort; as a result, such theories seem unpromising as accounts of genuinely responsible agency. We have just seen that one such account (Robert Nozick's) is clearly inadequate. I will turn now to our third, and final theorist, Robert Kane, who has offered a sophisticated alternative approach of the causal indeterminist variety that tries to improve on the evident shortcomings of its forebears.

V

Robert Kane's (1985) causal indeterminist theory (like Nozick's) suggests that the indeterminism that is necessary for human freedom "is to be located primarily in the will of the agent (in choice and decision) and only secondarily in overt action, through the relation of overt action to the will" (p.15). He defines choice as "the formation of intention, or the creation of (the state of having) a purpose or end" (p.17), and goes on to say that choices may, but need not, immediately give rise to action, and similarly, that actions may, but need not, be initiated by choices (p.18).

In attempting to come to terms with the problem I discussed in the previous section, Kane takes his cue from a suggestion made by David Wiggins:

...maybe all [libertarians] need to imagine or conceive is a world in which (a) there is some macroscopic indeterminacy founded in microscopic indeterminacy, and (b) an appreciable number of the free actions or policies or deliberations of individual agents, although they are not even in principle hypothetico-deductively derivable from antecedent conditions, can be such as to persuade us to fit them into meaningful sequences... it may not matter if the world approximates to a world which satisfies the principles of a neurophysiological determinism provided only that this fails in the last resort to characterize the world completely, and provided that there are actions which, for all that they are causally under-determined, are answerable to practical reason. (1973, pp. 52, 53)

Kane suggests that the libertarian does well to posit strict limitations on the extent of indeterminism in human behavior in two ways: the degree of indeterminacy within certain instances of (indeterministic) deliberative processes, and the range of choices in which indeterminism plays any role at all. The first of these limitations (as we will come to see)

plays a part in at least one of his attempts to show how an agent can bear responsibility for an undetermined choice.

He is led to the second of these limitations by his belief that any (credible) account of what the widely-presupposed capacity (in most circumstances) to have "chosen otherwise" comes to must entail that it is a power of "dual rational self control" - a "power to choose either way as the reasonable outcome of a deliberative process" (p.60, emphasis added). Since often the deliberative process quickly and directly points to one of the possible courses of action as being the most sensible one to choose, choosing any of the others in just those circumstances "would be arbitrary in relation to [this] prior psychological history" (p.57). But the standard incompatibilist analysis of our ability to choose otherwise is just to say that an alternative choice might have been made in precisely the same circumstances. Therefore, Kane concludes, given the constraint of dual rational self control, we cannot apply an incompatibilist analysis to our ascriptions of such an ability in circumstances of this (fairly common) sort. It is only in situations where we are uncertain about what is the best course of action to pursue that we might plausibly take ourselves to be free to choose and do otherwise than we in fact do in the "strong", indeterministic sense. (As we will see, he attempts to show that in this kind of circumstance an indeterministic ability to choose otherwise can satisfy the constraint he lays down.)

It won't be necessary to assess here the plausibility of either his constraint of "dual rational self control" or the conclusion that he draws from it concerning the extent of free, responsible action. I will move on instead to lay out his account of the nature of free choice. Kane develops his account in several stages, corresponding to different sorts of choice scenarios, each of which he regards as requiring separate analysis. I will confine myself to a discussion of two of these: "ordinary" practical choice and moral choice.

The cases falling under the heading of 'practical choice' are simply those deliberative situations in which duty/desire and short-term/long-term interest conflicts do not play a significant role.²⁶ Our representative example may be my having to choose which of two job offers to accept. Let us suppose that I am initially quite uncertain about which is preferable against the set of my relevant background beliefs and desires, but after a period of deliberation, I come to believe that one of the positions is clearly preferable to the other. Kane suggests that practical deliberation of this sort bears a significant amount of similarity to creative problem solving, in that both involve thought experiments in which random search procedures for new relevant considerations can aid the process. (He also draws a second analogy to the processes of mutation and natural selection in natural evolution.) Indeterminism or chance occurrences, then, may play the role in practical

choice of "initiating processes of thinking about or imagining, remembering or attending to, possible states or changes that may be relevant to deliberation in some way, as means or consequences of the options" (p.105).²⁷ Some such "considerings" are immediately rejected as irrelevant, but others are seen to be relevant and prompt the addition of a particular reason into the set of consciously recognized factors favoring the adoption of one of the options. (For example, while I am deliberating about which job offer to accept, I might recall "out of the blue" a visit to a friend on the West Coast, and in particular many memorable evening walks along the beach. This may serve either to remind or to make salient to me the positive effects that being near the ocean has consistently had on my sense of well-being.) If via this sort of process the agent comes to regard her total set of reasons as clearly favoring one of the alternatives and chooses accordingly, she will have made a rational decision in which indeterminism played a key role not in the final choice itself, but in the initial process of reasons-gathering leading up to the choice.

It seems to me, though, that it is unsatisfactory to suggest that it is only through the presence of random indeterministic mental occurrences that we may be significantly free with respect to the reflective process by which we come to consider new factors and ideas relevant to our practical choices. For, I would suggest, we want to say that we can, at times, actively direct our thoughts in a manner which is not completely determined.²⁸ But the possibility of thought processes which are both fully controlled and significantly free is not accounted for by Kane's view. Instead, indeterminism (or "chance", as he puts it) primarily has the negative role of "screen[ing] out all complete explanation, prediction, and control in terms of prior circumstances, so that the process can be a 'new beginning,' unfolding by virtue of its own internal rationality" (p.109). Perhaps we shall have to judge that Kane is right in supposing that there can be no thought processes that are controlled while undetermined, but if a theory can be constructed that seems to allow for this possibility, it would certainly be a preferable one for indeterminists.

I should note, however, that Kane does suggest that a more "active" sort of indeterministic process may be operative in many cases of practical deliberation. It is a kind of effort of will "to keep the doors of consciousness open to new or as yet unconsidered possibilities that might well up from the unconscious, from memory, or through free association, and might have a bearing on practical deliberation or [a creative] problem to be solved" (1988, p.458). This effort of will might itself be indeterministic.

While this suggestion does not attempt to meet the desiderata that I mention in the previous paragraph, it is helpful in pointing towards a somewhat different (yet common enough) sort of case: one in which I go through a mental effort of "rummaging" for further considerations that may bear on a decision I will eventually make or a problem I am trying

to solve. The metaphor of rummaging in a long-unvisited attic seems apt, since I have no sense of direct conscious control over the considerations that occur to me, but search about in various general directions in the hope that I will happen upon something useful. But if (as Kane intends) this sort of indeterministic effort to continue searching out further considerations is one for which it might be appropriate to commend an individual - and the lack of which could provide a legitimate basis (in some cases) for criticism - then surely we must be able to freely control it. If the causal indeterminist line that agent control obtains in virtue of an indeterministic causal connection is applied to such effort, then Kane must show how it can be a free type of control, how it can be up to the agent that this control is exercised in the particular way that it happens to be. Kane's most substantial response to this challenge is made in connection with his theory of moral (and prudential) choice, which has as a common (and prominent) feature the notion of indeterministic effort of will. Since what he says there can be readily adapted to the account of practical choice just given, I will simply proceed to this other account, and discuss Kane's response to the problem in that context.

Kane lays down the following conditions on an adequate libertarian theory of moral choice:

(UR) An agent has ultimate responsibility for the choice of A or some alternative (= it is "up to the agent here and now" in the sense of ultimate responsibility whether the agent chooses A or chooses otherwise), just in case:

(1) (The Production Condition) The choice that is made is the intentional termination of an effort of will that is the agent's effort of will.

(2) (The Rationality Condition) The agent (r₁) has reasons for so choosing, (r₂) does it for those reasons, (r₃) does not choose (for those reasons) compulsively, and (r₄) believes at the time of choice that the reasons for which it is made are in some sense the weightier reasons, more worth acting upon than their alternatives.

(3) (The Ultimacy Condition) Given the facts of the situation, no other explanation (other than the conjunction of (1) and (2)) for the agent's choice is possible, unless that explanation can in turn be explained by the conjunction of (1) and (2) itself. (I.e., the explanation provided by (1) and (2) is "ultimate" or "final".) In particular, any explanation of the agent's making the effort of will in (1) and of the agent's having the character and reasons or motives for choosing in (2) will not also explain the choice, even though (1) and (2) will explain the choice.²⁹

Just how we are to understand these conditions will become clearer below, but I should begin by spelling out the motivation behind each of them. (3) is a way of capturing the incompatibilist constraint that a choice for which an agent is responsible cannot be a deterministic product of factors obtaining prior to the agent's deliberation. (2) is an attempt to elucidate what is required for satisfaction of the generally-recognized condition that (at least paradigm) cases of free and responsible actions must be susceptible to explanation in terms of reasons. (2) and (3) together formalize the "dual rational self control" condition, noted earlier. Again, it is important to recognize that it is because Kane believes the latter to be a genuine constraint that he focuses on cases of apparent deliberative "struggle" as his paradigms of free choice. Finally, (1) is intended to provide the crucial component - and a sufficient condition, together with (2) and (3) - of an adequate response to the challenge we have posed to causal indeterminist theories generally, which Kane puts thus: "Indeterminism seems to undermine the idea that the agent is in control of the outcome whichever way it goes...[we may] call this the problem of dual production" (p.233).

I now will briefly state Kane's account of moral choices for which the agent is directly responsible. Once this account is before us, we will be able to understand how condition (1) (the condition of direct interest to us at the moment) is to be understood. I will then undertake to evaluate both his theory's success in satisfying (1) as well as the adequacy of this condition for a resolution to what Kane terms "the problem of dual production."

We may begin by noting that "efforts of will" in Kane's terminology are not "willings" in the traditional sense, but "mental efforts directed at getting one's ends (purposes, intentions) sorted out" (p.234). In a context of moral decision-making, they are struggles to act in accordance with perceived obligation in the face of contrary motives to act differently, with the outcome being uncertain in the agent's mind prior to the moment of choice. Now an explanation can readily be given for why such efforts occur prior to some choices but not others. It is that in cases requiring effort, one's prior character and motives provide strong, separate motivations for acting either way:

It is because the effort is thus a response to an inner conflict which is embedded in the prior character and motives that the prior character and motives can explain the conflict and explain why the effort is being made, without also explaining the outcome of the conflict and the effort. [They] provide reasons for going either

way, but not decisive reasons in the sense that would explain which way the agent would inevitably go. (p.235)

One's prior character and motives do not explain the choice which terminates the effort, Kane suggests, because that effort is itself an indeterminate process. Kane seems to conceive of this indeterminacy somewhat differently from Nozick. Instead of the weight of reasons being actually indeterminate prior to choice, they have values at any given time, but unstably so, fluctuating in an undetermined manner until the moment of choice.³⁰ As on Nozick's account, though, the choice brings it about that one set of reasons prevails, and issues in the corresponding action: "whichever way we choose (to overcome or to succumb), we will be making the reasons for which we choose at that moment the weightier ones for us, all things considered, by choosing" (p.242).

Finally, what is being underscored by pointing to the fact that an agent's choice is "the intentional termination of an effort of will that is the agent's effort of will," as the Production Condition requires? Simply that neither of the alternative motivations are compulsive in nature, for each stems from our own wills:

This is what makes these cases so important for understanding free will. If we fail to resist temptation it is because, though we wanted to resist, in another equally important sense we also wanted to succumb. Phenomenologically, we all recognize that these temptation cases are not like trying to hit a target and missing. In the temptation cases the resistance is coming from our own wills, and so the outcome is willed, whichever way it goes. (p.244)

We may now examine Kane's response to those who would challenge his claim that satisfaction of the Production Condition is sufficient (assuming that the Rationality and Ultimacy Conditions also have been met) for an agent's bearing ultimate responsibility for his choice - for its being "'up to the agent here and now'...whether the agent chooses A or chooses otherwise."

In his initial presentation of his views (1985), Kane suggested that the agent had only "partial" responsibility for his choice, in virtue of his having only "partial control" (pp.153-5). What Kane has in mind here is not very clear. First of all, one wants to know what Kane means by the term "partial control". He writes: "If you are going to reconcile the common intuitions that one controls one's free choices and that they are not determined, then you are going to have to talk about partial control. You cannot have it both ways." (p.155) It is hard to resist interpreting this and other, similar statements as

implying that control over one's choice is only partial in virtue of its not being causally determined. (And so 'control' in this context would be essentially my notion of 'agent-control'.³¹ I suggest another reading of Kane's use of this term immediately below.) But if that is his suggestion, then Kane seems to be guilty of the confusion noted by Bishop (discussed earlier): he fails to recognize that for the causal indeterminist, (agent-) control is constituted by a causal relation between reasons and choice (or action), and this is no "less" the case if the relation happens to be indeterministic. Tempting as such a reading is, though, it might be that an agent "controls" his action, in Kane's terminology, to the extent that it is up to him which action occurs. His concession, then, is that it is not entirely up to the agent which choice will occur, though, he insists, it is partly up to him (and so the agent is, accordingly, only "partly" responsible for it).³²

For this move to be successful, of course, we need to know why a causally undetermined decision is even partly up to the agent. And as we have noted, this clearly seems to require that it be (partly) up to the agent that one particular reason (or set of reasons) causally produced the decision, rather than some other reason that might have caused a different decision. Kane argues that this is the case because the link between the initial consideration of reasons and subsequent choice is mediated by an effort of will on the part of the agent. To the natural objection that this doesn't help matters since such effort is itself indeterminate on Kane's theory, he replies (p.154):

The degree of effort is indeed indeterminate. Yet if it is enough, then I make enough effort, and if it is not enough, then I do not make enough effort. I can kick myself for not having made more.... I cannot say that I made a certain (determinate) amount of effort and then it was a matter of luck as to whether that amount of effort was enough. One cannot say this because the amount of effort made was itself indeterminate. Similarly, one could not say of two persons in exactly the same situation, that they made the same amount of effort and, by chance or luck, one of them managed to overcome temptation while the other did not. Such claims are meaningless on [my] theory...The effort and the chance cannot be separated and viewed as independent variables. The chance is simply the indeterminateness of the effort. This is why it is never possible to absolve oneself from responsibility in moral choice situations on grounds of bad luck, even though the outcome is not determined.

What Kane is attempting to do here is to suggest that I am partly responsible for one action of mine (a decision) because its occurrence is causally influenced by another form

of activity in which I was engaged (effort of will), and for which I was responsible. But he seems to think that it is somehow inappropriate to inquire into the reason why I am responsible for this latter sort of activity, even though this kind of question can be quite sensibly asked concerning the former type. That is, it is not enough to say that my decision was up to me because it was caused but not necessitated by (one or more of) my reasons. For it must be the case that it was up to me - at least partly - that that reason would cause my decision in that particular case. We are then told that this requirement is met in virtue of the fact that the obtaining of that non-necessitating relation in this instance was influenced by an effort of will. As for this effort of will itself, well, it was my effort, wasn't it?

Clearly, this won't do, however. Kane can of course note that the fact that this effort of will occurred wasn't simply a random matter - it was the agent's own character and conflicting motives that induced the internal "struggle". But the particular way the effort of will turned out was not itself determined by his prior character and motives. It might have gone differently. So if the agent was responsible for how this period of deliberation shaped up, why it happened to go this way rather than some other way that might have significantly diminished the likelihood of the actual decision's obtaining, then it must have been up to him (or, if one likes, at least "partly" up to him) that those prior circumstances caused (while not necessitating) the effort of will to have that particular character rather than some other. We cannot adequately deal with this problem at one stage of activity (one's decision) simply by referring the questioner back to another (an "effort of will"). Since I do not believe Kane has any further resources to bring to bear on the question posed at this deeper level, I conclude that this attempt to resolve the problem fails.

Kane's response to this problem in the (1989) article abandons this tack of appealing to a kind of "partial" responsibility in favor of a more direct attempt to meet the challenge. He considers two different ways in which this challenge might be formulated. The first consists of asking the following question "Q": "Why did the agent intentionally terminate the effort of will at the time in this way rather than that, and for these reasons rather than those, given that the alternative choice might also have been rational?" (p.245) Kane claims that a straightforward explanation ("E") can be provided: "Because the agent came to believe at the time that these were all things considered the weightier reasons." But, of course, the agent's coming to believe this and his intentionally terminating a prior effort of will (i.e., choosing) are one and the same event, in Kane's view. Kane is aware of this, but urges that it is a peculiar feature of such events that they can be "self-explaining": "One can answer the question of why it occurred under one description by citing it under another description, and vice versa" (p.246). Now it is, perhaps,

informative to note that the agent's comparative weightings of his reasons at the time came about via his choice: it tells us at least the sort of event (a "volition" or choice) by which the reasons came to have their relative weights. But we must at once add two (closely connected) further points, each of which is in opposition to Kane's explanatory gambit: First, it seems that an explanatory relation must, of necessity, be antisymmetrical.³³ Kane fails to give any reason for limiting the scope of this truth, apart from saving the theory: "Only if acts are self explaining in this way, I believe, can regresses be avoided and Ultimacy secured." So even if the 'cognitive aspect' ("coming to believe") of an act could explain its obtaining under its 'volitional aspect' ("intentionally terminating"), it could not go both ways; the aspect that is the explanans would still need explanation (or have none).³⁴ Thus if acts are only partially self-explaining, then, contrary to what Kane suggests, it is not the case that regress can be avoided and ultimacy secured. But, secondly, it is clear that even partial (self-) explanation cannot be legitimately claimed, since a genuine explanation of those weightings must explain how it was produced. It is simply of no use to tell us that this particular kind of mental event has the peculiar feature of having distinct aspects to its intentional content that are internally linked in an interesting way. Given that an event cannot cause itself, in order to account for its occurrence we must refer to some factor that is logically distinct from it. So Kane is forced back to the reasons which caused, but did not necessitate, the choice. But, of course, this will not give us an account of why this choice rather than that was made, and so our problem remains. Again, the basic problem with Kane's analysis is that it assumes, without defense, the denial of a seemingly fundamental formal feature of explanatory relations — this time, that they are irreflexive.³⁵

Part of what drives Kane to such desperate measures is his belief that for any choice for which an agent S can be truly responsible, it must be the case that (a) there are at least two mutually incompatible options available to S, the choice of either of which could be a reasonable outcome of a process of deliberation relative to his total set of reasons just prior to that choice (i.e., neither has the preponderance of support, so that choosing the other alternative would be "arbitrary", rationally speaking); and yet (b) there ultimately must be an explanation of why one choice rather than the other was made, in terms of reasons alone. These constraints are simply not jointly satisfiable. I believe that neither of these is in fact a necessary condition on responsible choice, though I will not undertake a defense of this claim here. We may note, however, that Kane does make some remarks that are independent of the unsatisfactory suggestion above, and that also appear to address (though in a quite different manner) our question as to whether it is up to the agent which of a pair of possible choices will actually occur, given that it is causally

possible that either of them be caused (but not determined) by a particular set of prior factors, and that there are no other causally relevant factors beyond those potential causes.

He suggests (p.248) that we must distinguish two kinds of control over one's choices - "absolute or complete" and "ultimate". The latter variety "requires that actions be undetermined by events or occurrences." In raising our question, Kane claims, we can only be hankering after a sort of control ("absolute control") in virtue of which "the choice was determined by a determinate effort of the agent." But this is for the choice to be necessitated by prior factors. So obviously, one can't have both. (It should be clear that these types of control correspond to the indeterministic and deterministic causal relation theories of agent-control introduced earlier.) Kane is not explicit about how these points are to be linked to the question at hand, but I think his view can be reconstructed from his various remarks as follows:

The concept of moral responsibility for one's actions is a coherent one. The concept requires that whether or not an agent performs an action must be in some sense "up to him". This can only mean that he exercises one or the other of the two possible forms of control over his action. By familiar arguments, moral responsibility is not compatible with causal determinism. Therefore, the sort of control by which the occurrence of a free action is up to the agent must be the one that is granted under indeterminism, which is "ultimate control".

This argument constitutes an entirely defensive sort of maneuver.³⁶ It appeals to the fact that few of us would wish to concede that morally responsible behavior is impossible. Kane acknowledges that he is "uneasy" about the sort of theory it requires him to defend, but believes it to be the only route available to the incompatibilist.³⁷ The problem is that the question we have been posing for his causal indeterminist theory - How is it up to an agent that his prior inner state (indeterministically) caused this choice rather than any of the others that were also causally possible outcomes of that state? - seems to be a proper and important one. Moreover, most incompatibilists acknowledge its force. Now many incompatibilists are led by this consideration to espouse some version of the agency theory. Kane believes that this sort of view cannot be coherently made out, and so thinks we are pretty much stuck with the less-than-satisfying causal indeterminacy view.

If he is right, however, a possibility that must be seriously considered is that our conception of morally responsible action is conceptually tied to a vague (but ultimately

incoherent) conception of agent causality, and so must be abandoned with it.³⁸ I have not attempted to settle this question here. What I have tried to do, though, is to provide some grounds for thinking that the available alternatives (for the libertarian theorist) to the agency theory are not particularly promising.³⁹

Bibliography

- Anscombe, G.E.M. (1971). Causality and Determination. Cambridge: University Press.
- Bishop, John (1989). Natural Agency. Cambridge: University Press.
- Dennett, Daniel (1978). "On Giving Libertarians What They Say They Want", in Brainstorms. Cambridge, MA: Bradford Books.
- Double, Richard (1991). The Non-Reality of Free Will. New York: Oxford University Press.
- Ginet, Carl (1980). "The Conditional Analysis of Freedom", in Peter van Inwagen, ed., Time and Cause. Dordrecht: D.Reidel, 171-86.
- Honderich, Ted (1988). A Theory of Determinism. Oxford: University Press.
- Kane, Robert (1985). Free Will and Values. Buffalo: SUNY Press.
- (1989). "Two Kinds of Incompatibilism", Philosophy and Phenomenological Research 50, 219-54.
- Kenny, Anthony (1968). Descartes. New York: Random House.
- Nagel, Thomas (1986). The View From Nowhere. New York: Oxford University Press.
- Nozick, Robert (1981). Philosophical Explanations. Cambridge: Harvard University Press.
- O'Connor, Timothy (1992). "Some Puzzles About Free Agency". Ph.D. dissertation, Cornell University.
- (forthcoming). "On the Transfer of Necessity", Nous.
- Sorabji, Richard (1980). Necessity, Cause, and Blame. Ithaca: Cornell University Press.
- Strawson, Galen (1986). Freedom and Belief. Oxford: University Press.
- Taylor, Richard (1966). Action and Purpose. Englewood Cliffs: Prentice Hall.
- van Inwagen, Peter (1983). An Essay on Free Will. Oxford: University Press.
- Williams, Bernard (1978). Descartes: The Project of Pure Enquiry. New York: Penguin Books.

¹ Formal modal versions of this argument may be found in Ginet (1980) and van Inwagen (1983). For a discussion of various critical reactions to their argument, see my "On the Transfer of Necessity", forthcoming in *Nous*.

² See, e.g., Nagel (1986), Strawson (1986), and Double (1991).

³ Because I want to focus on the central issue of how these theories account for the possibility of responsibility-conferring control over one's actions within an indeterministic framework, I will not be addressing directly other important issues in the general theory of action. So, e.g., I will not attempt to settle whether a libertarian theorist should posit something that might properly be termed "volition" (and if so, what account should be given of the nature of volition), how this may or may not be related to an agent's choices or decisions, and whether or not every free action is at least partly constituted by some sort of decision. I will assume that at least many of our actions involve movement-triggering decisions (conceived of as the agent's coming to have a type of intention). I will also steer clear of fundamental controversies in the philosophy of mind. Although certain accounts that I will be looking at are cast in terms of particular theories of the nature of mental states, the general issues I will be concerned with are not (at least not obviously) affected by the adoption of those theories rather than other standard approaches.

⁴ At least when those requirements are given fairly abstract characterizations (as I have done above in sketching the notion of agent-control).

⁵ I am not suggesting that these are the only two conditions on responsible agency; rather, they are simply the only ones that concern me here.

⁶ Two of the three theories I look at are variations on the same basic approach.

⁷ I have made a start on this task in "Some Puzzles About Free Agency" (Ph.D. dissertation, Cornell University, 1992). I hope to publish the relevant material in the near future.

⁸ The usage of this phrase that is in view here is not the one which is perhaps more common, viz., giving an occurrent (or about-to-occur) event a certain characteristic (such as my spoiling your party by offending one of the guests). The discussion in the text is concerned, rather, with the notion of initiating an event (or the process leading to that event).

⁹ Much more, of course, would have to be said here in an attempt to give a complete compatibilist characterization of the notion of "control" (e.g., explaining how one's control is thought to be augmented through the development of a stable character).

¹⁰ An influential argument for the claim that the concept of causation (whatever is typically involved in actual causal processes) does not require necessitation can be found in Anscombe (1971). Her argument is endorsed in, e.g., Sorabji (1982), Ch.2 and van Inwagen (1983), Ch.4. An opponent of this contention is Honderich (1988) (see Ch.1 and pp.201-2).

¹¹ pp.144ff. Though van Inwagen is committed to rejecting this argument, he readily admits that he is unable to show what is wrong with it. (Indeed, he says that he finds it "puzzling" that the proposition (S), given immediately below, should be false. But for reasons that I won't spell out here, he takes himself to be confronted with a choice between the merely "puzzling" and the "inconceivable", and so he opts for the former.)

¹² Bishop does not, however, advocate an indeterministic theory of human agency. His purpose in defending the indeterminist on this point lies in his belief that the thrust of the considerations noted in the text are really directed against any form of causal theory of action, indeterministic or otherwise. I think he is mistaken about this, as will be implicit in my criticism of his discussion below.

¹³ Perhaps it is often more probable that a particular action will be causally produced in virtue of the strength of the reason suggesting it.

¹⁴ Van Inwagen does not distinguish between the issue of agent-control and that of whether or not one has a genuine choice about the action one performs, and so leaves it open for the reader to take him to understand these to be one and the same, as Bishop does. I think it is clear, however, that when van Inwagen speaks about "having a choice about" whether an event occurs, he has a stronger notion in mind than what I am terming basic "agent-control". (Like so many other terms from popular usage that are taken over for philosophical purposes, "control" no doubt has slightly varying connotations when used by different philosophers. For some, the instantiation of the concept they have in mind may be sufficient for freedom of choice. I think it is helpful to separate these terms, however, since the term "agent-control" seems particularly apt to capture the common ground between compatibilists and some libertarians concerning an important criterion for free and responsible agency. And developing a conceptual framework for zeroing in on the precise points of disagreement (which are often inchoate or only dimly articulated) helps to advance the discussion on this, the most intractable of philosophical problems.)

¹⁵ And, I would argue, that is implicit in the practice of practical deliberation.

¹⁶ p.239, n.34.

¹⁷ Indeed, Mark Crimmins has responded in this way, after reading a previous draft of this paper.

¹⁸ Our present desires and beliefs generally not being under our immediate control.

¹⁹ It is essentially the same principle as Ginet (1980)'s "modus ponens for power necessity" and van Inwagen (1983)'s "B-rule". For reasons that are irrelevant to the example in the text, one needs to be careful about the way this principle is formulated. See my "On the Transfer of Necessity", cited above.

²⁰ Nozick (1981), pp.291-316. Like van Inwagen's argument in the previous section, Nozick's discussion comes with a disclaimer: He agrees with the incompatibilist that there is a value that we attach to our actions that is incompatible with causal determinism. (He terms this "originative value". It would be an unnecessary digression to explain this here, but his discussion of this notion (pp.310ff.) is an illuminating account of what the incompatibilist wants to secure in repudiating universal causal determinism.) But he is doubtful that indeterminism of the requisite sort obtains, and unsure of whether a coherent account of responsible yet indeterministic agency can be given. So the account given in the text is a highly tentative proposal, one about which Nozick himself is suspicious. I will try to show that his suspicions are well-grounded.

²¹ Non-self-subsuming decisions "that are based on weights previously given in such decisions, revokable weights, will inherit autonomy" (p.301). Nozick does not attempt to elucidate just how this inheritance of autonomy might work. Perhaps what he has in mind is this: since such decisions reflect the previously-chosen conception of oneself, they are subsumed by, and so (partly) explained by, this previous act. Yet, wouldn't they also be self-subsuming insofar as they appear to involve a free reaffirmation of this conception?

²² Note that this is not to ask: "For what reasons did the agent assign those weights (rather than some others)?" Since the very purpose of the agent's decision is to rank his preferences or values, there may be no answer to this question.

²³ I'm referring not to the Fifth Meditation, but to a passage in the First Reply:

First, possible existence, at the very least, attends to such a being....Next, when we attend to the immense power of this being, we shall be unable to think of its existence as possible without also recognizing that it can exist by its own power; and we shall infer from this that this being does really exist and has existed from eternity, since it is quite evident by the natural light that what can exist by its own power always exists. (AT VII,119; p.85 of the Cottingham, Stoothoff, & Murdoch translation)

Anthony Kenny and Bernard Williams each see Descartes as implicitly endorsing here the premise that everything desires existence (and so God, being capable of existing by his own might, can will his own existence). (See Kenny (1968), pp.159-64 and Williams (1978), pp.157-60.)

²⁴ I should caution the reader that I certainly don't mean to give the impression that I (like Nozick) believe that a genuinely explanatory connection must involve subsumption under a law.

²⁵ In practice, this often will not be as trivial as it sounds: a person's subsequent affirmation of a decision that marked a restructuring of values may come from those priorities having become deeply entrenched over time, whereas the initial adoption of them was somewhat tentative and experimental.

²⁶ Situations in which these types of considerations are paramount Kane terms "moral" and "prudential", respectively. Though I shall consider only the first of these explicitly, they are given parallel analyses by Kane and are, accordingly, subject to the same liabilities.

²⁷ As Kane notes, a similar suggestion to the one in the text had been made independently by Daniel Dennett. (See Dennett (1978), especially pp.293ff.)

²⁸ See Taylor (1966), pp.58-9 and 156-8, for a brief characterization of the phenomenological difference between "active" and "passive" thought processes.

²⁹ Kane (1989), pp.231-2. (I have slightly modified Kane's statement of (UR) for the sake of greater clarity.) The references throughout the remainder of my discussion of Kane will be to this work. His account here is similar to that given in Kane (1985), but I think it is more clearly presented in the later work.

³⁰ See Kane (1985), pp.146-9.

³¹ I, of course, do not accept Kane's claim that agent-control can only be "partial" in the absence of causal determination.

³² In correspondence, Prof. Kane has informed me that the latter of these readings was, in fact, the intended one.

³³ A relation R is antisymmetrical iff, for any x and y, if Rxy and Ryx, then it necessarily follows that x=y.

³⁴ Thus, in claiming that Kane's theory violates the antisymmetry of explanation, I am taking it that the "cognitive aspect" and "volitional aspect" of a decision/volition are to be treated as distinct entities when analyzing the formal properties of

the relation Kane believes to obtain, since neither is said to provide an explanation for the decision/volition simpliciter.

³⁵ It will be observed that each of the objections I have raised are instances of the more general point that explanatory relations are asymmetrical. (R is asymmetrical iff R is antisymmetrical and irreflexive.) I should note that Kane clearly recognizes the requirement of asymmetry in relation to the causal relation, and is denying only that it is true (without exception) of the broader relation of explanation. But, as I argue in the text, Kane fails to motivate the exclusion of instances of free will from such a requirement on explanation. Furthermore, it strikes me as odd for the causal indeterminist, who recognizes that an agent's control over her decision is realized only through the appropriate instantiation of a causal relation, to assert that, nonetheless, the explanation of why this choice rather than that was made is to be found in certain intrinsic features of the choice itself.

³⁶ In broad outline, it is similar to the route taken by van Inwagen.

³⁷ We might recall in this connection Kane's suggestion in his earlier work that the agent's responsibility might in some sense be only "partial". Though he wisely drops this from his later discussion, the theory does not substantially change, thus giving further indication of Kane's awareness that his causal indeterminist approach is less than wholly satisfactory.

³⁸ As I noted at the outset of this paper, some recent authors have drawn just this conclusion. See, e.g., Strawson (1986), Ch.2, and Nagel (1986), Ch.7.

³⁹ I wish to thank Carl Ginet, Norman Kretzmann, Mark Crimmins, and two anonymous reviewers of this journal for helpful comments on an earlier draft of this work. I also thank Robert Kane for general discussion of the relevant issues in correspondence as well as detailed comments on a previous draft of this paper.